Adaptive inexact semismooth Newton methods for the contact problem between two membranes*

Jad Dabaghi [†‡] Vincent Martin [§] Martin Vohralík[†‡]

April 28, 2020

Abstract

We propose an adaptive inexact version of a class of semismooth Newton methods that is aware of the continuous (variational) level. As a model problem, we study the system of variational inequalities describing the contact between two membranes. This problem is discretized with conforming finite elements of order $p \geq 1$, yielding a nonlinear algebraic system of variational inequalities. We consider any iterative semismooth linearization algorithm like the Newton-min or the Newton–Fischer–Burmeister which we complement by any iterative linear algebraic solver. We then derive an a posteriori estimate on the error between the exact solution at the continuous level and the approximate solution which is valid at any step of the linearization and algebraic resolutions. Our estimate is based on flux reconstructions in discrete subspaces of $\mathbf{H}(\mathrm{div},\Omega)$ and on potential reconstructions in discrete subspaces of $H^1(\Omega)$ satisfying the constraints. It distinguishes the discretization, linearization, and algebraic components of the error. Consequently, we can formulate adaptive stopping criteria for both solvers, giving rise to an adaptive version of the considered inexact semismooth Newton algorithm. Under these criteria, the efficiency of the leading estimates is also established, meaning that we prove them equivalent with the error up to a generic constant. Numerical experiments for the Newton-min algorithm in combination with the GMRES algebraic solver confirm the efficiency of the developed adaptive method.

Keywords: variational inequality, complementarity condition, contact problem, semismooth Newton method, a posteriori error estimate, adaptivity, stopping criterion

1 Introduction

Consider a system of algebraic inequalities written in the following form: find a vector $\boldsymbol{X}_h \in \mathbb{R}^n$, such that

$$\begin{aligned} \mathbb{E}\boldsymbol{X}_h &= \boldsymbol{F}, \\ \boldsymbol{K}(\boldsymbol{X}_h) &\geq \boldsymbol{0}, \quad \boldsymbol{G}(\boldsymbol{X}_h) \geq \boldsymbol{0}, \quad \boldsymbol{K}(\boldsymbol{X}_h) \cdot \boldsymbol{G}(\boldsymbol{X}_h) = 0, \end{aligned} \tag{1}$$

where, for some integers $n > 1$ and $0 < m < n$, $\mathbb{E} \in \mathbb{R}^{n-m,n}$ is a matrix, $\boldsymbol{K} : \mathbb{R}^n \to \mathbb{R}^m$ and $\boldsymbol{G} : \mathbb{R}^n \to \mathbb{R}^m$ are affine operators, and $\boldsymbol{F} \in \mathbb{R}^{n-m}$ is a given vector. The first line of (1) typically represents the discretization of a linear partial differential equation (PDE) (the model example for this study is described further in (6)). The second line of (1) represents linear complementarity constraints and states that the vectors $\boldsymbol{K}(\boldsymbol{X}_h)$ and $\boldsymbol{G}(\boldsymbol{X}_h)$ have non-negative components and are orthogonal. Numerous algorithms have been developed in the past for the approximate solution of (1), see for example the overview of Facchinei and Pang [26, 27] and the books of Bonnans, Gilbert, Lemaréchal, and Sagastizábal [8], Ito and Kunisch [33], and Ulbrich [50]. In particular, we mention the approach by interior point method of Wright [54], the active set strategy by

*This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 647134 GATIPOR).

†Inria, 2 rue Simone Iff, 75589 Paris, France

‡Université Paris-Est, CERMICS (ENPC), 77455 Marne-la-Vallée 2, France

§Université technologie de Compiègne (UTC), 60200, France

Kanzow [36], and the primal-dual active set strategy, closely linked to semismooth Newton methods, see Hintermüler, Ito, and Kunisch [29, 32]. Alternatively, in [34, 48, 40, 49, 31, 51], a sequence of regularized problems is solved, coupled to a path-following strategy to choose the associated parameter.

The approach that we use here is to rewrite directly the complementarity conditions in the second line of (1) as a system of nonsmooth nonlinear equations by means of C-functions, see [20, 26, 27, 7]. The C-functions are not smooth in the classical sense (Fréchet-differentiable), but admit a weaker smoothness (the Clarke derivative), *cf.* [17]. This yields an equivalent formulation of (1) that requests to find a vector $\boldsymbol{X}_h \in \mathbb{R}^n$ such that

$$\mathcal{S}(\boldsymbol{X}_h) = \boldsymbol{0}, \tag{2}$$

where $\mathcal{S} : \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear non-differentiable function. Next, let any semismooth nonlinear solver be applied to system (2), yielding at step $k \geq 1$ a linear system

$$\mathbb{A}^{k-1}\boldsymbol{X}_h^k = \boldsymbol{B}^{k-1}, \tag{3}$$

where $\mathbb{A}^{k-1} \in \mathbb{R}^{n,n}$ is a matrix and $\boldsymbol{B}^{k-1} \in \mathbb{R}^n$ is a vector. Finally, let any iterative algebraic solver be applied to (3), yielding at step $i \geq 1$ an approximation $\boldsymbol{X}_h^{k,i}$ to $\boldsymbol{X}_h$. Note that $\boldsymbol{X}_h^{k,i}$ does not solve (3) but only

$$\mathbb{A}^{k-1}\boldsymbol{X}_h^{k,i} = \boldsymbol{B}^{k-1} - \boldsymbol{R}^{k,i}, \tag{4}$$

where $\boldsymbol{R}^{k,i} := \boldsymbol{B}^{k-1} - \mathbb{A}^{k-1}\boldsymbol{X}_h^{k,i} \in \mathbb{R}^n$ is the algebraic residual vector of (3). Similarly, $\boldsymbol{X}_h^{k,i}$ does not solve (2) as $\mathcal{S}(\boldsymbol{X}_h^{k,i}) \neq 0$ in general. Our first goal is to perform an a posteriori analysis of problem (1), where the matrix $\mathbb{E}$ is given by a discretization of the underlying PDE. We are namely interested in deriving a fully computable upper bound on the energy error $e(\boldsymbol{X}_h^{k,i})$ between the approximate solution associated with the algebraic vector $\boldsymbol{X}_h^{k,i}$ and the unknown solution of the continuous-level variational inequality in the form

$$e(\boldsymbol{X}_h^{k,i}) \leq \eta(\boldsymbol{X}_h^{k,i}) = \eta_{\text{disc}}^{k,i} + \eta_{\text{lin}}^{k,i} + \eta_{\text{alg}}^{k,i}. \tag{5}$$

Here, the a posteriori error estimate $\eta(\boldsymbol{X}_h^{k,i})$ is fully computable from $\boldsymbol{X}_h^{k,i}$ at each step $k \geq 1$, $i \geq 1$. As our second goal, we distinguish in $\eta(\boldsymbol{X}_h^{k,i})$ the components of the error caused by the discretization, the linearization, and the algebraic resolution. Finally, our third goal is to conceive an adaptive inexact algorithm based on a posteriori stopping criteria. These request to stop the algebraic (respectively linearization) solver whenever the algebraic estimator $\eta_{\text{alg}}^{k,i}$ (respectively the linearization estimator $\eta_{\text{lin}}^{k,i}$) does not contribute significantly to the overall estimator $\eta(\boldsymbol{X}_h^{k,i})$. We thus propose an answer the two following practical questions: 1) To which precision should (3) and (2) be solved? 2) What is the error in $\boldsymbol{X}_h^{k,i}$?

Our general viewpoint is that if one uses a semismooth Newton method (4), then the adaptive inexact algorithm based on the estimates (5) may bring an important computational speed-up, in addition to the fact that the overall error can be assessed at any moment. Actually, the proposed a posteriori error analysis, aware of the PDE level, may steer the semismooth Newton method rather differently than what is usual. For instance, in our approach, one may not reach at all the region of the fast (quadratic/superlinear) convergence of the semismooth Newton method, since the total error is dominated by the discretization error component and our adaptive algorithm stops the semismooth Newton iterations prior to entering the fast convergence zone, see, e.g., the right plot in Figure 5 below. Note also that we do not employ here any regularization.

An important amount of work has been performed in the last years on a posteriori analysis of partial differential equations (see for instance the books of Verfürth [53], Ainsworth [1] and Repin [46] for a general introduction). Concerning a posteriori error estimates for variational inequalities discretized as in (1) or (2), let us mention the pioneering work of Brezzi, Hager, and Raviart [13], next Ainsworth, Oden, and Lee [2], Kornhuber [39], Repin [47] and Bürg and Schröder [15]. For the elliptic obstacle problem we can more precisely mention the papers of Veeser [52], Chen and Nochetto [16], and Braess [9]. Not to solve (3) exactly or with a high precision leads to the concept of an inexact semismooth Newton method. Such approaches are heavily used in practice and theoretical foundations can be found in [14, 22, 38] for the case of inexact Newton methods and in [25, 26, 27, 37, 43, 28] for inexact semismooth Newton methods. All these approaches, however, do not take into account the discretization error of the PDE by the given numerical scheme, only addressing the convergence of $\boldsymbol{X}_h^{k,i}$ to $\boldsymbol{X}_h$ in the above example, whereas we rather steer our algorithm by the estimated distance of $\boldsymbol{X}_h^{k,i}$ to the PDE solution $\boldsymbol{X}$. The general concepts we use

to derive (5) follow Becker, Johnson, and Rannacher [3], Louf, Combe, and Pelle [42], Jiránek *et al.* [35], Ern and Vohralík [23], and Papež *et al.* [45, 44]. In particular, to achieve a guaranteed bound of the form (5), we use the equilibrated flux reconstructions with auxiliary local problems by Destuynder and Métivet [21] and Braess and Schöberl [11]. A reconstruction of the primal variable satisfying the constraints on the given step $k \geq 1$, $i \geq 1$, will also be performed.

Let $\Omega \subset \mathbb{R}^2$ be a polygon. We exemplify the above approach with the following problem that models the contact between two membranes: find u_1, u_2, and λ such that

$$\left\{\begin{array}{llll} -\mu_1 \Delta u_1 - \lambda = f_1 & \text{in} & \Omega, & \\ -\mu_2 \Delta u_2 + \lambda = f_2 & \text{in} & \Omega, & \\ u_1 - u_2 \geq 0, \quad \lambda \geq 0, \quad (u_1 - u_2)\lambda = 0 & \text{in} & \Omega, & \\ u_1 = g & \text{on} & \partial\Omega, & \\ u_2 = 0 & \text{on} & \partial\Omega, & \end{array}\right. \tag{6}$$

where u_1 and u_2 represent vertical displacements of the two membranes and λ is a Lagrange multiplier characterizing the action of the second membrane on the first one. The constant parameters $\mu_1, \mu_2 > 0$ correspond to the tension of the membranes, and $f_1, f_2 \in L^2(\Omega)$ are given external forces. The boundary condition prescribed by a constant $g > 0$ ensures that the first membrane is above the second one on the boundary $\partial\Omega$. In (6), the two first equations represent the kinematic behavior of the membranes, and the third one represents the linear complementarity conditions saying that either the membranes are separated ($u_1 > u_2$, $\lambda = 0$), or they are in contact ($u_1 = u_2$, $\lambda \geq 0$). [—] A combined path-following semismooth Newton strategy for problem (6) at the continuous level was recently proposed and analyzed in Zhang, Yan, and Ran [55]. A finite element discretization together with an a priori convergence analysis was performed in [4, 5], and an a posteriori analysis was undertaken in [6]. Therein, however, it was supposed that the discrete system (1) is solved exactly, for continuous and piecewise affine finite elements. The additional difficulty we have to treat here is that our approximate solutions do not fulfill the constraints (because of the inexact solve (4) for any polynomial degree $p \geq 1$, and in general for $p \geq 2$).

This contribution is organized as follows. In Section 2, the model problem (6) is discretized by finite elements of any polynomial degree $p \geq 1$, yielding an algebraic system of the form (1) [—]. In Section 3, we present the concept of the inexact semismooth Newton method giving rise to systems (2)–(4). The various flux reconstructions [—] are described in Section 4. Next, Section 5 is dedicated to the construction of the a posteriori error estimate of the form (5). In Section 6, we present the adaptive inexact semismooth algorithm and in Section 7, we prove the converse inequality to (5) (up to a generic constant) for the leading terms, assessing the quality of our estimates. Finally, Section 8 is devoted to numerical experiments for $p = 1$ and $p = 2$.

2 Model problem and its finite element discretization

In this section, we set up the notation, describe in details the model problem (6), and introduce its finite element discretization for all polynomial degrees $p \geq 1$. For the sake of brevity, the results in Sections 2.3–2.5 are given without proofs which can be found in [18, Sect.1.2].

2.1 Function spaces and basic notation

Let $H^1(\Omega)$ be the space of L^2 functions on the domain Ω which admit a weak gradient in $[L^2(\Omega)]^2$ and $H^1_0(\Omega)$ its zero-trace subspace. Similarly, $\mathbf{H}(\mathrm{div}, \Omega)$ stands for the space of $[L^2(\Omega)]^2$ functions having a weak divergence in $L^2(\Omega)$. Moreover, we define the set $H^1_g(\Omega) := \big\{v \in H^1(\Omega),\ v = g \text{ on } \partial\Omega\big\}$. The standard notations $\boldsymbol{\nabla}$ and $\boldsymbol{\nabla}\cdot$ are used respectively for the weak gradient and divergence operators. For a nonempty set $\mathcal{O}$ of $\mathbb{R}^2$, we denote its Lebesgue measure by $|\mathcal{O}|$ and the $L^2(\mathcal{O})$ scalar product by $(u,v)_{\mathcal{O}} := \int_{\mathcal{O}} uv \,\mathrm{d}x$ for $u, v \in L^2(\mathcal{O})$. We also use the following notations: $\|v\|^2_{\mathcal{O}} := (v,v)_{\mathcal{O}}$ and $\|\boldsymbol{\nabla} v\|^2_{\mathcal{O}} := (\boldsymbol{\nabla} v, \boldsymbol{\nabla} v)_{\mathcal{O}}$; when $\mathcal{O} = \Omega$, the index is dropped. Besides, the Poincaré–Friedrichs and the Poincaré–Wirtinger inequalities state that if $\overline{v}_{\mathcal{O}}$ denotes the mean value of v on $\mathcal{O}$ and $h_{\mathcal{O}}$ the diameter of $\mathcal{O}$, then

$$\|v\|_{\mathcal{O}} \leq C_{\mathrm{PF}} h_{\mathcal{O}} \|\boldsymbol{\nabla} v\|_{\mathcal{O}} \qquad \forall v \in H^1_0(\mathcal{O}), \tag{7a}$$

$$\|v - \overline{v}_{\mathcal{O}}\|_{\mathcal{O}} \leq C_{\mathrm{PW}} h_{\mathcal{O}} \|\boldsymbol{\nabla} v\|_{\mathcal{O}} \qquad \forall v \in H^1(\mathcal{O}). \tag{7b}$$

The constants $C_{\rm PF}$ and $C_{\rm PW}$ can be precisely estimated in many cases. In particular, $C_{\rm PF}$ is at most 1 and if $\mathcal{O}$ is convex, then $C_{\rm PW}$ can be taken as $\frac{1}{\pi}$. We define the energy norm

$$|||\boldsymbol{v}|||_{\mathcal{O}} := \left\{ \sum_{\alpha=1}^{2} \mu_\alpha \left\| \boldsymbol{\nabla} v_\alpha \right\|_{\mathcal{O}}^2 \right\}^{\frac{1}{2}}, \qquad \boldsymbol{v} = (v_1, v_2) \in [H_0^1(\mathcal{O})]^2. \tag{8}$$

When $\mathcal{O} = \Omega$, we use the shorthand notation $|||\boldsymbol{v}||| := |||\boldsymbol{v}|||_{\mathcal{O}}$. We also define the following rescaling of the $H^{-1}(\mathcal{O})$ norm:

$$\forall v \in H^{-1}(\mathcal{O}), \quad |||v|||_{H_*^{-1}(\mathcal{O})} := \sup_{\psi \in H_0^1(\mathcal{O}),\ \max(\mu_1^{\frac{1}{2}}, \mu_2^{\frac{1}{2}}) \|\boldsymbol{\nabla}\psi\|_{\mathcal{O}} = 1} \langle v, \psi \rangle . \tag{9}$$

2.2 Full and reduced problems

Set $\boldsymbol{u} := (u_1, u_2)$, $\boldsymbol{v} := (v_1, v_2)$ and define the forms

$$a(\boldsymbol{u}, \boldsymbol{v}) := \sum_{\alpha=1}^{2} \mu_\alpha \left(\boldsymbol{\nabla} u_\alpha, \boldsymbol{\nabla} v_\alpha \right), \ b(\boldsymbol{v}, \chi) := \left(\chi, v_1 - v_2 \right), \ l(\boldsymbol{v}) := \sum_{\alpha=1}^{2} \left(f_\alpha, v_\alpha \right); \tag{10}$$

note that a is coercive on $[H_0^1(\Omega)]^2$. Let us also define the convex set

$$\Lambda := \left\{ \chi \in L^2(\Omega),\ \chi \geq 0 \text{ a.e. in } \Omega \right\}.$$

Supposing $(f_1, f_2) \in [L^2(\Omega)]^2$ and g a positive constant, the weak formulation of (6) consists in finding $\boldsymbol{u} \in H_g^1(\Omega) \times H_0^1(\Omega)$ and $\lambda \in \Lambda$ such that

$$a(\boldsymbol{u}, \boldsymbol{v}) - b(\boldsymbol{v}, \lambda) = l(\boldsymbol{v}) \quad \forall \boldsymbol{v} \in [H_0^1(\Omega)]^2, \tag{11a}$$

$$b\left(\boldsymbol{u}, \chi - \lambda\right) \geq 0 \qquad \forall \chi \in \Lambda. \tag{11b}$$

Following [5, Proposition 1], (11) admits a unique weak solution. Define also the convex set $\boldsymbol{\mathcal{K}}_g$ by

$$\boldsymbol{\mathcal{K}}_g := \left\{ (v_1, v_2) \in H_g^1(\Omega) \times H_0^1(\Omega),\ v_1 - v_2 \geq 0 \text{ a.e. in } \Omega \right\}. \tag{12}$$

Then a reduced variational problem, equivalent to (11) (*cf.* [5, Lemma 2]) is to find $\boldsymbol{u} = (u_1, u_2) \in \boldsymbol{\mathcal{K}}_g$ such that

$$a(\boldsymbol{u}, \boldsymbol{v} - \boldsymbol{u}) \geq l(\boldsymbol{v} - \boldsymbol{u}) \qquad \forall \boldsymbol{v} = (v_1, v_2) \in \ \boldsymbol{\mathcal{K}}_g. \tag{13}$$

(13) is classically well-posed, cf. Lions and Stampacchia [41] or Hlaváček *et al.* [30].

2.3 Discretization of the reduced problem by finite elements

Let $\mathcal{T}_h$ be a conforming simplicial mesh of Ω, *i.e.* $\mathcal{T}_h$ is a set of triangles verifying $\cup_{K \in \mathcal{T}_h} \overline{K} = \overline{\Omega}$, where the intersection of the closure of two elements of $\mathcal{T}_h$ is either an empty set, a vertex, or an edge. The set of vertices of $\mathcal{T}_h$ is denoted by $\mathcal{V}_h$ and is partitioned into the interior vertices $\mathcal{V}_h^{\rm i}$ and the boundary vertices $\mathcal{V}_h^{\rm e}$. The vertices of an element $K \in \mathcal{T}_h$ are collected in the set $\mathcal{V}_K$. Denote by h_K the diameter of a triangle K and $h := \max_{K \in \mathcal{T}_h} h_K$. Furthermore, for a vertex $\boldsymbol{a} \in \mathcal{V}_h$, let the patch $\omega_h^{\boldsymbol{a}} \subset \Omega$ be the domain made up of the elements of $\mathcal{T}_h$ that share $\boldsymbol{a}$. The vector $\boldsymbol{n}_{\omega_h^{\boldsymbol{a}}}$ stands for its outward unit normal.

In the sequel, we use the discrete conforming space of piecewise polynomial functions

$$X_h^p := \left\{ v_h \in \mathcal{C}^0(\overline{\Omega});\ v_h|_K \in \mathbb{P}_p(K) \ \ \forall K \in \mathcal{T}_h \right\} \subset H^1(\Omega),$$

where $\mathbb{P}_p(K)$ stands for the set of polynomials of total degree less than or equal to p on the element $K \in \mathcal{T}_h$. We consider any $p \geq 1$. We also denote by $\mathcal{V}^p$ the set of the Lagrange nodes $\boldsymbol{x}_l$ and by $\mathcal{N}^p$ its cardinality. The interior nodes are collected in the set $\mathcal{V}^{p,{\rm i}}$ (with $\mathcal{N}^{p,{\rm i}}$ its cardinality)and the boundary ones are collected in the set $\mathcal{V}^{p,{\rm e}}$. The Lagrange basis functions of X_h^p are then denoted by $(\psi_{h,\boldsymbol{x}_l})_{1 \leq l \leq \mathcal{N}^p}$, $\boldsymbol{x}_l \in \mathcal{V}^p$; $\psi_{h,\boldsymbol{x}_l}$ takes value one in $\boldsymbol{x}_l$ and zero in all other Lagrange nodes. In the particular case $p = 1$, the set $\mathcal{V}^1$ coincides

with the mesh vertices $\mathcal{V}_h$, and the Lagrange basis functions are the "hat" basis functions denoted by $\psi_{h,\boldsymbol{a}}$, $\boldsymbol{a} \in \mathcal{V}_h$.

We also introduce the boundary-aware set and space

$$X^p_{gh} := \{v_h \in X^p_h,\ v_h = g \text{ on } \partial\Omega\} \subset H^1_g(\Omega), \qquad X^p_{0h} := X^p_h \cap H^1_0(\Omega),$$

as well as the convex set where the constrains are only imposed in the Lagrange nodes

$$\boldsymbol{\mathcal{K}}^p_{gh} := \left\{ \boldsymbol{v}_h = (v_{1h}, v_{2h}) \in X^p_{gh} \times X^p_{0h},\ v_{1h}(\boldsymbol{x}_l) - v_{2h}(\boldsymbol{x}_l) \geq 0 \ \ \forall \boldsymbol{x}_l \in \mathcal{V}^{p,\mathrm{i}} \right\}. \tag{14}$$

Recall (12) and observe that $\boldsymbol{\mathcal{K}}^1_{gh} \subset \boldsymbol{\mathcal{K}}_g$ holds but $\boldsymbol{\mathcal{K}}^p_{gh} \not\subset \boldsymbol{\mathcal{K}}_g$ when $p \geq 2$. The discrete counterpart to (13) then consists in finding $\boldsymbol{u}_h = (u_{1h}, u_{2h}) \in \boldsymbol{\mathcal{K}}^p_{gh}$ such that

$$a(\boldsymbol{u}_h, \boldsymbol{v}_h - \boldsymbol{u}_h) \geq l(\boldsymbol{v}_h - \boldsymbol{u}_h) \qquad \forall \boldsymbol{v}_h = (v_{1h}, v_{2h}) \in \boldsymbol{\mathcal{K}}^p_{gh}. \tag{15}$$

As a result of the Lions–Stampacchia theorem, problem (15) admits a unique solution.

Following the methodology of [5, equation (4.5)] or [15], let for all $(w_h, v_h) \in X^p_h \times X^p_h$

$$\langle w_h, v_h \rangle_h := \begin{cases} \displaystyle\sum_{\boldsymbol{a} \in \mathcal{V}_h} w_h(\boldsymbol{a}) v_h(\boldsymbol{a}) \frac{|\omega^{\boldsymbol{a}}_h|}{3} & \text{if} \quad p = 1, \qquad (16a) \\ (w_h, v_h) & \text{if} \quad p \geq 2. \qquad (16b) \end{cases}$$

Then we define a discrete convex set

$$\Lambda^p_h := \left\{ v_h \in X^p_h;\ \langle v_h, \psi_{h,\boldsymbol{x}_l} \rangle_h \geq 0 \ \forall \boldsymbol{x}_l \in \mathcal{V}^{p,\mathrm{i}},\ \langle v_h, \psi_{h,\boldsymbol{x}_l} \rangle_h = 0 \ \forall \boldsymbol{x}_l \in \mathcal{V}^{p,\mathrm{e}} \right\}. \tag{17}$$

Observe that $\Lambda^p_h \not\subset \Lambda$ for $p \geq 2$, whereas in the case $p = 1$, Λ^p_h reduces to

$$\Lambda^1_h = \left\{ v_h \in X^1_{0h};\ v_h(\boldsymbol{a}) \geq 0 \quad \forall \boldsymbol{a} \in \mathcal{V}^{\mathrm{i}}_h \right\} = \left\{ v_h \in X^1_{0h};\ v_h \geq 0 \right\} \subset \Lambda, \tag{18}$$

same as in [4, Section 4]. The sets $\boldsymbol{\mathcal{K}}^p_{gh}$ and Λ^p_h are chosen to satisfy the following property that will give rise to a discrete weak formulation for the constraints:

$$\langle \chi_h, v_{1h} - v_{2h} \rangle_h = \sum_{l=1}^{\mathcal{N}^{p,\mathrm{i}}} (v_{1h} - v_{2h})(\boldsymbol{x}_l) \langle \chi_h, \psi_{h,\boldsymbol{x}_l} \rangle_h \geq 0 \quad \forall \chi_h \in \Lambda^p_h,\ \forall (v_{1h}, v_{2h}) \in \boldsymbol{\mathcal{K}}^p_{gh}.$$

Finally, let λ_{1h} and λ_{2h} in X^p_h be given by

$$\begin{array}{rcll} \langle \lambda_{1h}, \psi_{h,\boldsymbol{x}_l} \rangle_h &=& \mu_1 (\boldsymbol{\nabla} u_{1h}, \boldsymbol{\nabla} \psi_{h,\boldsymbol{x}_l}) - (f_1, \psi_{h,\boldsymbol{x}_l}) & \forall \boldsymbol{x}_l \in \mathcal{V}^{p,\mathrm{i}}, \\ \langle \lambda_{1h}, \psi_{h,\boldsymbol{x}_l} \rangle_h &=& 0 & \forall \boldsymbol{x}_l \in \mathcal{V}^{p,\mathrm{e}}, \\ \langle \lambda_{2h}, \psi_{h,\boldsymbol{x}_l} \rangle_h &=& -\mu_2 (\boldsymbol{\nabla} u_{2h}, \boldsymbol{\nabla} \psi_{h,\boldsymbol{x}_l}) + (f_2, \psi_{h,\boldsymbol{x}_l}) & \forall \boldsymbol{x}_l \in \mathcal{V}^{p,\mathrm{i}}, \\ \langle \lambda_{2h}, \psi_{h,\boldsymbol{x}_l} \rangle_h &=& 0 & \forall \boldsymbol{x}_l \in \mathcal{V}^{p,\mathrm{e}}. \end{array} \tag{19}$$

Note that (16a) corresponds to the use of a mass lumping, so that (19) for $p = 1$ is a local postprocess, whereas for $p \geq 2$, the mass matrices in (19) are not diagonal. Extending [4, Proposition 12] to the case $p \geq 2$, we can easily obtain:

Lemma 2.1. *Let $(u_{1h}, u_{2h}) \in \boldsymbol{\mathcal{K}}^p_{gh}$ be the solution of the reduced discrete problem (15). Then the functions λ_{1h} and λ_{2h} defined by (19) coincide, we can set $\lambda_h := \lambda_{1h} = \lambda_{2h}$, and there holds $\lambda_h \in \Lambda^p_h$.*

[—]

2.4 Equivalence with the discretization of the full problem by finite elements

Relying on the (discrete when $p = 1$) L^2 scalar product of (16), one can also consider a discretization of (11) as: find $\boldsymbol{u}_h = (u_{1h}, u_{2h}) \in X^p_{gh} \times X^p_{0h}$ and $\lambda_h \in \Lambda^p_h$ such that

$$a(\boldsymbol{u}_h, \boldsymbol{v}_h) - \langle \lambda_h, v_{1h} - v_{2h} \rangle_h = l(\boldsymbol{v}_h) \quad \forall \boldsymbol{v}_h = (v_{1h}, v_{2h}) \in [X^p_{0h}]^2, \tag{20a}$$

$$\langle \chi_h - \lambda_h, u_{1h} - u_{2h} \rangle_h \geq 0 \qquad \forall \chi_h \in \Lambda^p_h. \tag{20b}$$

In extension of [5, Lemma 13] to $p \geq 2$, it can be seen that the equivalence from the continuous level carries over to the discrete one, see [18, Lemma 1.2.4] for the proof:

Lemma 2.2. *For any solution $(u_{1h}, u_{2h}, \lambda_h)$ of problem (20), the pair (u_{1h}, u_{2h}) is a solution of problem (15). Conversely, for any solution (u_{1h}, u_{2h}) of problem (15), defining the function $\lambda_h = \lambda_{\alpha h}$, $\alpha = 1, 2$ by (19), the triple $(u_{1h}, u_{2h}, \lambda_h)$ is a solution of problem (20).*

[—]

Remark 2.3. *Consider $\chi_h = 0$ and $\chi_h = 2\lambda_h \in \Lambda_h^p$ in (20b). Combining this with the definitions (17) of Λ_h^p and (14) of $\boldsymbol{\mathcal{K}}_{gh}^p$ gives the discrete complementarity constraints*

$$\begin{aligned} &(u_{1h} - u_{2h})(\boldsymbol{x}_l) \geq 0, \quad \langle \lambda_h, \psi_{h,\boldsymbol{x}_l} \rangle_h \geq 0 \quad \forall \boldsymbol{x}_l \in \mathcal{V}^{p,\mathrm{i}}, \quad \langle \lambda_h, \psi_{h,\boldsymbol{x}_l} \rangle_h = 0 \quad \forall \boldsymbol{x}_l \in \mathcal{V}^{p,\mathrm{e}}, \\ &\langle \lambda_h, u_{1h} - u_{2h} \rangle_h = 0. \end{aligned} \tag{21}$$

Note that when piecewise affine finite elements are employed ($p = 1$), (21) reduces to

$$(u_{1h} - u_{2h})(\boldsymbol{a}) \geq 0, \quad \lambda_h(\boldsymbol{a}) \geq 0, \quad \lambda_h(\boldsymbol{a})(u_{1h} - u_{2h})(\boldsymbol{a}) = 0 \quad \forall \boldsymbol{a} \in \mathcal{V}_h^{\mathrm{i}}, \tag{22}$$

so that in particular

$$u_{1h} \geq u_{2h}, \quad \lambda_h \geq 0 \qquad \text{if } p = 1, \tag{23}$$

and the approximation is conforming in that $\boldsymbol{u}_h \in \boldsymbol{\mathcal{K}}_g$ and $\lambda_h \in \Lambda$; more precisely, the two first equations of the constraints in (6) hold strongly (everywhere) for $(u_{1h}, u_{2h}, \lambda_h)$ when $p = 1$, whereas the third one is only satisfied discretely in the interior vertices. For $p \geq 2$, the approximation is generally nonconforming with $\boldsymbol{u}_h \notin \boldsymbol{\mathcal{K}}_g$, $\lambda_h \notin \Lambda$, and with L^2 integral product being zero only in place of the third constraint in (6).

2.5 Algebraic formulation as a complementarity problem

In order to express the discrete problem (20) under an algebraic form, consider the basis $(\Theta_{h,\boldsymbol{x}_l})_{1 \leq l \leq \mathcal{N}^p}$ of X_h^p, dual to $(\psi_{h,\boldsymbol{x}_l})_{1 \leq l \leq \mathcal{N}^p}$ in that

$$\begin{aligned} \langle \Theta_{h,\boldsymbol{x}_l}, \psi_{h,\boldsymbol{x}_l} \rangle_h &= 1 \qquad \forall \boldsymbol{x}_l \in \mathcal{V}^p, \\ \langle \Theta_{h,\boldsymbol{x}_l}, \psi_{h,\boldsymbol{x}_m} \rangle_h &= 0 \qquad \forall \boldsymbol{x}_l, \boldsymbol{x}_m \in \mathcal{V}^p, \boldsymbol{x}_m \neq \boldsymbol{x}_l. \end{aligned} \tag{24}$$

Note that for $p = 1$, the dual basis is just the Lagrange basis $\psi_{h,\boldsymbol{a}}$, $\boldsymbol{a} \in \mathcal{V}_h$, with the scaling $3/|\omega_h^{\boldsymbol{a}}|$, whereas for $p \geq 2$, each $\Theta_{h,\boldsymbol{x}_l}$ can be determined by inverting the finite element mass matrix as in (19). An important property is that $\Theta_{h,\boldsymbol{x}_l}$ belong to Λ_h^p for all $\boldsymbol{x}_l \in \mathcal{V}^{p,\mathrm{i}}$.

Noting that X_{gh}^p is decomposed as $X_{gh}^p = X_{0h}^p + g$ (recall that $g > 0$ is constant), and using (21), (24) for $\lambda_h \in \Lambda_h^p$, the three unknowns live in the internal nodes, *i.e.*

$$u_{1h} = \sum_{l=1}^{\mathcal{N}^{p,\mathrm{i}}} (\boldsymbol{X}_{1h})_l \psi_{h,\boldsymbol{x}_l} + g, \quad u_{2h} = \sum_{l=1}^{\mathcal{N}^{p,\mathrm{i}}} (\boldsymbol{X}_{2h})_l \psi_{h,\boldsymbol{x}_l}, \quad \lambda_h = \sum_{l=1}^{\mathcal{N}^{p,\mathrm{i}}} (\boldsymbol{X}_{3h})_l \Theta_{h,\boldsymbol{x}_l}. \tag{25}$$

Consequently, (20a) gives rise to a system of linear equations $\mathbb{E}\boldsymbol{X}_h = \boldsymbol{F}$, where $\boldsymbol{X}_h^T := (\boldsymbol{X}_{1h}, \boldsymbol{X}_{2h}, \boldsymbol{X}_{3h})^T \in \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}}}$ and the rectangular matrix $\mathbb{E} \in \mathbb{R}^{2\mathcal{N}^{p,\mathrm{i}}, 3\mathcal{N}^{p,\mathrm{i}}}$ is defined by

$$\mathbb{E} := \begin{bmatrix} \mu_1 \mathbb{S} & \mathbf{0} & -\mathbb{I}_{\mathrm{d}} \\ \mathbf{0} & \mu_2 \mathbb{S} & \mathbb{I}_{\mathrm{d}} \end{bmatrix},$$

with $\mathbb{S} \in \mathbb{R}^{\mathcal{N}^{p,\mathrm{i}}, \mathcal{N}^{p,\mathrm{i}}}$ the finite element stiffness matrix, $\mathbb{S}_{l,m} := (\boldsymbol{\nabla}\psi_{h,\boldsymbol{x}_m}, \boldsymbol{\nabla}\psi_{h,\boldsymbol{x}_l})$, $1 \leq l, m \leq \mathcal{N}^{p,\mathrm{i}}$, and $\mathbb{I}_{\mathrm{d}} \in \mathbb{R}^{\mathcal{N}^{p,\mathrm{i}}, \mathcal{N}^{p,\mathrm{i}}}$ the identity matrix. The right-hand side $\boldsymbol{F}$ is defined by blocks $\boldsymbol{F}^T := (\boldsymbol{F}_1, \boldsymbol{F}_2)^T$ with $(\boldsymbol{F}_\alpha)_l := (f_\alpha, \psi_{h,\boldsymbol{x}_l})$, $1 \leq l \leq \mathcal{N}^{p,\mathrm{i}}$, $\alpha = 1, 2$. Problem (20), taking into account (21) and relying on (24), can then be written under the compact form: find $\boldsymbol{X}_h \in \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}}}$ such that

$$\begin{aligned} \mathbb{E}\boldsymbol{X}_h &= \boldsymbol{F}, \\ \boldsymbol{X}_{1h} + g\mathbf{1} - \boldsymbol{X}_{2h} &\geq \mathbf{0}, \quad \boldsymbol{X}_{3h} \geq \mathbf{0}, \quad (\boldsymbol{X}_{1h} + g\mathbf{1} - \boldsymbol{X}_{2h}) \cdot \boldsymbol{X}_{3h} = 0, \end{aligned} \tag{26}$$

where $\mathbf{1} = [1, \ldots, 1]^T \in \mathbb{R}^{\mathcal{N}^{p,\mathrm{i}}}$. Consequently, denoting $\boldsymbol{K}(\boldsymbol{X}_h) := \boldsymbol{X}_{1h} + g\mathbf{1} - \boldsymbol{X}_{2h}$ and $\boldsymbol{G}(\boldsymbol{X}_h) := \boldsymbol{X}_{3h}$, which are respectively affine and linear, (26) fits the abstract class of problems (1) of the introduction.

2.6 C-functions

We now express the complementarity constraints in (26), which take a form of inequalities, as non-differentiable equalities. Let us recall that a function $f : (\mathbb{R}^m)^2 \to \mathbb{R}^m$, $m \geq 1$, is a C-function or a complementarity function if

$$\forall (\boldsymbol{x}, \boldsymbol{y}) \in (\mathbb{R}^m)^2 \qquad f(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{0} \quad \Longleftrightarrow \quad \boldsymbol{x} \geq \boldsymbol{0}, \quad \boldsymbol{y} \geq \boldsymbol{0}, \quad \boldsymbol{x} \cdot \boldsymbol{y} = 0.$$

Examples of C-functions are respectively the min function, the Fischer–Burmeister function, or the Mangasarian function

$$(\min\{\boldsymbol{x}, \boldsymbol{y}\})_l := \min\{\boldsymbol{x}_l, \boldsymbol{y}_l\} \qquad l = 1, \ldots, m, \tag{27a}$$

$$(f_{\mathrm{FB}}(\boldsymbol{x}, \boldsymbol{y}))_l := \sqrt{\boldsymbol{x}_l^2 + \boldsymbol{y}_l^2} - (\boldsymbol{x}_l + \boldsymbol{y}_l) \quad l = 1, \ldots, m, \tag{27b}$$

$$(f_{\mathrm{M}}(\boldsymbol{x}, \boldsymbol{y}))_l := \xi(|\boldsymbol{x}_l - \boldsymbol{y}_l|) - \xi(\boldsymbol{y}_l) - \xi(\boldsymbol{x}_l) \quad l = 1, \ldots, m, \tag{27c}$$

where $\xi : \mathbb{R} \mapsto \mathbb{R}$ is an increasing function satisfying $\xi(0) = 0$. For more details on C-functions see [26, 27]. Let $\tilde{\mathbf{C}}$ be any C-function, *i.e.*, satisfying, for $m = \mathcal{N}^{p,\mathrm{i}}$, $\tilde{\mathbf{C}}(\boldsymbol{X}_{1h} + g\mathbf{1} - \boldsymbol{X}_{2h}, \boldsymbol{X}_{3h}) = 0 \iff \boldsymbol{X}_{1h} + g\mathbf{1} - \boldsymbol{X}_{2h} \geq 0$, $\boldsymbol{X}_{3h} \geq 0$, and $(\boldsymbol{X}_{1h} + g\mathbf{1} - \boldsymbol{X}_{2h}) \cdot \boldsymbol{X}_{3h} = 0$. Then, introducing the function $\mathbf{C} : \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}}} \to \mathbb{R}^{\mathcal{N}^{p,\mathrm{i}}}$ defined as $\mathbf{C}(\boldsymbol{X}_h) := \tilde{\mathbf{C}}(\boldsymbol{X}_{1h} + g\mathbf{1} - \boldsymbol{X}_{2h}, \boldsymbol{X}_{3h})$, problem (26) can be equivalently rewritten as: find $\boldsymbol{X}_h \in \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}}}$ such that

$$\left\{ \begin{array}{ll} \mathbb{E}\boldsymbol{X}_h & = \boldsymbol{F}, \\ \mathbf{C}(\boldsymbol{X}_h) & = \boldsymbol{0}. \end{array} \right. \tag{28}$$

3 Inexact semismooth Newton methods

We now consider an approximation of the discrete system (26), rewritten using any C-function as a system of nonlinear algebraic equations (28), by a semismooth Newton method.

3.1 A semismooth Newton linearization

Given an initial vector $\boldsymbol{X}_h^0 \in \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}}}$, on step $k \geq 1$, one looks for $\boldsymbol{X}_h^k \in \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}}}$ such that

$$\mathbb{A}^{k-1} \boldsymbol{X}_h^k = \boldsymbol{B}^{k-1}, \tag{29}$$

where the Jacobian matrix $\mathbb{A}^{k-1} \in \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}}, 3\mathcal{N}^{p,\mathrm{i}}}$ and the right-hand-side vector $\boldsymbol{B}^{k-1} \in \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}}}$ are respectively defined by

$$\mathbb{A}^{k-1} := \left[\begin{array}{c} \mathbb{E} \\ \mathbf{J}_{\mathbf{C}}(\boldsymbol{X}_h^{k-1}) \end{array} \right], \qquad \boldsymbol{B}^{k-1} := \left[\begin{array}{c} \boldsymbol{F} \\ \mathbf{J}_{\mathbf{C}}(\boldsymbol{X}_h^{k-1}) \boldsymbol{X}_h^{k-1} - \mathbf{C}(\boldsymbol{X}_h^{k-1}) \end{array} \right]. \tag{30}$$

Note that since the first line of (28) is linear, the corresponding Jacobian is constant and equal to $\mathbb{E}$. The semismooth nonlinearity occurs in the second line of (28), so that $\mathbf{J}_{\mathbf{C}}(\boldsymbol{X}_h^{k-1})$ is the Clarke subdifferential of the semismooth C-function $\mathbf{C}$ at $\boldsymbol{X}_h^{k-1}$, see [8, 26, 27].

3.2 Example of a semismooth Newton method: the min case $p = 1$

For the semismooth function min (27a), we in particular obtain

$$\min\{\boldsymbol{X}_{1h} + g\mathbf{1} - \boldsymbol{X}_{2h}, \boldsymbol{X}_{3h}\} = \min\left\{ \left(\begin{array}{c} u_{1h}(\boldsymbol{x}_1) - u_{2h}(\boldsymbol{x}_1) \\ \vdots \\ u_{1h}(\boldsymbol{x}_{\mathcal{N}^{p,\mathrm{i}}}) - u_{2h}(\boldsymbol{x}_{\mathcal{N}^{p,\mathrm{i}}}) \end{array} \right), \left(\begin{array}{c} (\boldsymbol{X}_{3h})_1 \\ \vdots \\ (\boldsymbol{X}_{3h})_{\mathcal{N}^{p,\mathrm{i}}} \end{array} \right) \right\}.$$

If the block matrices $\mathbb{K}$ and $\mathbb{G}$ in $\mathbb{R}^{\mathcal{N}^{p,\mathrm{i}}, 3\mathcal{N}^{p,\mathrm{i}}}$ are defined by $\mathbb{K} := [\mathbb{I}_{\mathrm{d}}, -\mathbb{I}_{\mathrm{d}}, \mathbf{0}]$, $\mathbb{G} := [\mathbf{0}, \mathbf{0}, \mathbb{I}_{\mathrm{d}}]$, then the l^{th} row of the Jacobian matrix $\mathbf{J}_{\mathbf{C}}(\boldsymbol{X}_h^{k-1})$ is either given by the l^{th} row of $\mathbb{K}$ if $u_{1h}^{k-1}(\boldsymbol{x}_l) - u_{2h}^{k-1}(\boldsymbol{x}_l) \leq \left(\boldsymbol{X}_{3h}^{k-1}\right)_l$, or by the l^{th} row of $\mathbb{G}$ if $u_{1h}^{k-1}(\boldsymbol{x}_l) - u_{2h}^{k-1}(\boldsymbol{x}_l) > \left(\boldsymbol{X}_{3h}^{k-1}\right)_l$.

3.3 Inexact solution of the linear algebraic systems (general case $p \geq 1$)

As a crucial point in our study, we focus on the case where the system of linear algebraic equations (29) is solved inexactly. Suppose thus that some iterative algebraic solver is applied to the linear system (29). Given an initial vector $\boldsymbol{X}_h^{k,0} \in \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}}}$, often taken as $\boldsymbol{X}_h^{k,0} = \boldsymbol{X}_h^{k-1}$, this yields on step $i \geq 1$ an approximation $\boldsymbol{X}_h^{k,i}$ to $\boldsymbol{X}_h^k$ satisfying

$$\mathbb{A}^{k-1}\boldsymbol{X}_h^{k,i} = \boldsymbol{B}^{k-1} - \boldsymbol{R}_h^{k,i}, \tag{31}$$

where $\boldsymbol{R}_h^{k,i} := \boldsymbol{B}^{k-1} - \mathbb{A}^{k-1}\boldsymbol{X}_h^{k,i} \in \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}}}$ is the algebraic residual vector. Note that $\boldsymbol{R}_h^{k,i}$ has a block structure of the form $(\boldsymbol{R}_h^{k,i})^T := (\boldsymbol{R}_{1h}^{k,i}, \boldsymbol{R}_{2h}^{k,i}, \boldsymbol{R}_{3h}^{k,i})^T$, with $\boldsymbol{R}_{1h}^{k,i} \in \mathcal{N}^{p,\mathrm{i}}$ corresponds to the test functions v_{1h} in (20a) (with $v_{2h} = 0$), $\boldsymbol{R}_{2h}^{k,i}$ corresponds to the test functions v_{2h} in (20a) (with $v_{1h} = 0$), and $\boldsymbol{R}_{3h}^{k,i}$ issues from the complementarity constraints (21). The approximations $(u_{1h}^{k,i}, u_{2h}^{k,i}, \lambda_h^{k,i})$ are then obtained from $\boldsymbol{X}_h^{k,i}$ as in (25).

4 Flux reconstructions

We introduce here flux reconstructions that will be central in our a posteriori analysis. We follow some general concepts in [11, 21, 23] and the references therein. Let $k \geq 1$ be a semismooth linearization step and $i \geq 1$ be a linear solver step. Denote by $\boldsymbol{\Pi}_{\mathbb{P}_p}$ the L^2-orthogonal projection onto the space $\mathbb{P}_p(\mathcal{T}_h)$ of discontinuous piecewise polynomials of order $p \geq 1$. We in particular construct here $\boldsymbol{\sigma}_{\alpha h}^{k,i} \in \mathbf{H}(\mathrm{div}, \Omega)$, $\alpha \in \{1, 2\}$, such that

$$\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}_{\alpha h}^{k,i} = \boldsymbol{\Pi}_{\mathbb{P}_p}(f_\alpha) - (-1)^\alpha \lambda_h^{k,i} \quad \in \mathbb{P}_p(\mathcal{T}_h), \tag{32}$$

i.e.,

$$\left(\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}_{\alpha h}^{k,i} + (-1)^\alpha \lambda_h^{k,i}, q_h\right)_K = (f_\alpha, q_h)_K \quad \forall q_h \in \mathbb{P}_p(K),\ \forall K \in \mathcal{T}_h.$$

The construction of these fluxes is based on the first two diffusion equations in (6) that are linear. Consequently, we do not need to construct any linearization fluxes as in [23]. The fluxes $\boldsymbol{\sigma}_{\alpha h}^{k,i}$ are an approximation in $\mathbf{H}(\mathrm{div}, \Omega)$ to the opposite of the gradient of $u_{\alpha h}^{k,i}$ multiplied by μ_α. We will further separate them into two contributions: one lifting the algebraic residuals $\boldsymbol{R}_{1h}^{k,i}$ and $\boldsymbol{R}_{2h}^{k,i}$ of Section 3.3 and the other dealing with the discretization error. [—]

4.1 Algebraic residual representation

Following [45, 44], we first associate with $\boldsymbol{R}_{1h}^{k,i}$ and $\boldsymbol{R}_{2h}^{k,i}$ of Section 3.3 discontinuous piecewise polynomials $r_{1h}^{k,i}$ and $r_{2h}^{k,i}$ of degree $p \geq 1$ that vanish on the boundary of Ω. These can be easily computed solving on each element $K \in \mathcal{T}_h$ a small problem with mass matrix as follows. For $\boldsymbol{x}_l \in \mathcal{V}^{p,\mathrm{i}}$, denote by $N_{h,\boldsymbol{x}_l}$ the number of mesh elements forming the support of the basis function $\psi_{h,\boldsymbol{x}_l}$. Then, $\forall K \in \mathcal{T}_h$, $\forall \alpha \in \{1, 2\}$, define $r_{\alpha h}^{k,i}|_K \in \mathbb{P}_p(K)$ by

$$(r_{\alpha h}^{k,i}, \psi_{h,\boldsymbol{x}_l})_K = \frac{(\boldsymbol{R}_{\alpha h}^{k,i})_l}{N_{h,\boldsymbol{x}_l}}, \quad r_{\alpha h}^{k,i}|_{\partial K \cap \partial\Omega} := 0$$

for all basis functions $\psi_{h,\boldsymbol{x}_l}, \boldsymbol{x}_l \in \mathcal{V}^{p,\mathrm{i}}$ nonzero on K. It is easily seen that the first $2\mathcal{N}^{p,\mathrm{i}}$ lines of (31) then read, *cf.* (20a) and (19),

$$\begin{aligned}
\mu_1 \left(\boldsymbol{\nabla} u_{1h}^{k,i}, \boldsymbol{\nabla}\psi_{h,\boldsymbol{x}_l}\right) &= \left(f_1 + \tilde{\lambda}_{h,l}^{k,i} - r_{1h}^{k,i}, \psi_{h,\boldsymbol{x}_l}\right) \quad \forall l = 1, \ldots, \mathcal{N}^{p,\mathrm{i}}, \\
\mu_2 \left(\boldsymbol{\nabla} u_{2h}^{k,i}, \boldsymbol{\nabla}\psi_{h,\boldsymbol{x}_l}\right) &= \left(f_2 - \tilde{\lambda}_{h,l}^{k,i} - r_{2h}^{k,i}, \psi_{h,\boldsymbol{x}_l}\right) \quad \forall l = 1, \ldots, \mathcal{N}^{p,\mathrm{i}},
\end{aligned} \tag{33}$$

where

$$\tilde{\lambda}_{h,l}^{k,i} := \begin{cases} \lambda_h^{k,i}(\boldsymbol{x}_l) \text{ (real number given by the vertex value of } \lambda_h^{k,i}) \text{ if } p = 1, \\ \lambda_h^{k,i} \text{ (function } \lambda_h^{k,i}, \text{ the index } l \text{ being discarded) if } p \geq 2. \end{cases} \tag{34}$$

In the sequel, we also use the shorthand notation, for a vertex $\boldsymbol{a} \in \mathcal{V}_h$,

$$\tilde{\lambda}_{h,\boldsymbol{a}}^{k,i} := \begin{cases} \lambda_h^{k,i}(\boldsymbol{a}) & \text{if } p = 1, \\ \lambda_h^{k,i} & \text{if } p \geq 2. \end{cases} \tag{35}$$

4.2 Discretization flux reconstruction

We now provide a way to obtain the discretization flux reconstructions $(\boldsymbol{\sigma}_{1h,\mathrm{disc}}^{k,i}, \boldsymbol{\sigma}_{2h,\mathrm{disc}}^{k,i})$. This is done via solution of local mixed systems on the patches $\omega_h^{\boldsymbol{a}}$ around the mesh vertices $\boldsymbol{a} \in \mathcal{V}_h$ of the mesh $\mathcal{T}_h$ and crucially employs the $\mathbb{P}_1$ hat basis functions $\psi_{h,\boldsymbol{a}}$ that form a partition of unity by $\sum_{\boldsymbol{a}\in\mathcal{V}_h} \psi_{h,\boldsymbol{a}} = 1$. The Raviart–Thomas spaces of order $p \geq 1$ [12] are defined by

$$\mathbf{RT}_p(\Omega) := \{\boldsymbol{\tau}_h \in \mathbf{H}(\mathrm{div},\Omega),\ \boldsymbol{\tau}_h|_K \in \mathbf{RT}_p(K)\ \ \forall K \in \mathcal{T}_h\},$$

where $\mathbf{RT}_p(K) := [\mathbb{P}_p(K)]^2 + \vec{\boldsymbol{x}}\mathbb{P}_p(K)$, with $\vec{\boldsymbol{x}} = [x_1, x_2]^T$. For a vertex $\boldsymbol{a} \in \mathcal{V}_h$, let

$$\mathbf{RT}_p(\omega_h^{\boldsymbol{a}}) := \{\boldsymbol{\tau}_h \in \mathbf{H}(\mathrm{div},\omega_h^{\boldsymbol{a}}),\ \boldsymbol{\tau}_h|_K \in \mathbf{RT}_p(K),\ \forall K \in \mathcal{T}_h \text{ such that } K \subset \omega_h^{\boldsymbol{a}}\},$$

and let $\mathbb{P}_p(\mathcal{T}_h|_{\omega_h^{\boldsymbol{a}}})$ stand for piecewise discontinuous polynomials of order $p \geq 1$ in the patch $\omega_h^{\boldsymbol{a}}$. Define consequently the spaces $\mathbf{V}_h^{\boldsymbol{a}}$ and $Q_h^{\boldsymbol{a}}$, when $\boldsymbol{a} \in \mathcal{V}_h^{\mathrm{i}}$, by

$$\mathbf{V}_h^{\boldsymbol{a}} := \left\{\boldsymbol{\tau}_h \in \mathbf{RT}_p(\omega_h^{\boldsymbol{a}}),\ \boldsymbol{\tau}_h\cdot\boldsymbol{n}_{\omega_h^{\boldsymbol{a}}} = 0 \text{ on } \partial\omega_h^{\boldsymbol{a}}\right\},\quad Q_h^{\boldsymbol{a}} := \left\{q_h \in \mathbb{P}_p(\mathcal{T}_h|_{\omega_h^{\boldsymbol{a}}}),\ (q_h,1)_{\omega_h^{\boldsymbol{a}}} = 0\right\}$$

and, when $\boldsymbol{a} \in \mathcal{V}_h^{\mathrm{e}}$, by

$$\mathbf{V}_h^{\boldsymbol{a}} := \left\{\boldsymbol{\tau}_h \in \mathbf{RT}_p(\omega_h^{\boldsymbol{a}}),\ \boldsymbol{\tau}_h\cdot\boldsymbol{n}_{\omega_h^{\boldsymbol{a}}} = 0 \ \text{ on } \ \partial\omega_h^{\boldsymbol{a}}\backslash\partial\Omega\right\},\quad Q_h^{\boldsymbol{a}} := \mathbb{P}_p(\mathcal{T}_h|_{\omega_h^{\boldsymbol{a}}}).$$

Definition 4.1. *Let $\left(u_{1h}^{k,i}, u_{2h}^{k,i}, \lambda_h^{k,i}\right)$ be the approximate solution given by (31), verifying in particular (33). For each vertex $\boldsymbol{a} \in \mathcal{V}_h$, define $\boldsymbol{\sigma}_{\alpha h,\mathrm{disc}}^{k,i,\boldsymbol{a}} \in \mathbf{V}_h^{\boldsymbol{a}}$ and $\gamma_{\alpha h}^{k,i,\boldsymbol{a}} \in Q_h^{\boldsymbol{a}}$, by solving:*

$$\begin{aligned}
\left(\boldsymbol{\sigma}_{\alpha h,\mathrm{disc}}^{k,i,\boldsymbol{a}}, \boldsymbol{\tau}_h\right)_{\omega_h^{\boldsymbol{a}}} - \left(\gamma_{\alpha h}^{k,i,\boldsymbol{a}}, \boldsymbol{\nabla}\cdot\boldsymbol{\tau}_h\right)_{\omega_h^{\boldsymbol{a}}} &= -\left(\mu_\alpha\psi_{h,\boldsymbol{a}}\boldsymbol{\nabla}u_{\alpha h}^{k,i}, \boldsymbol{\tau}_h\right)_{\omega_h^{\boldsymbol{a}}} && \forall\boldsymbol{\tau}_h \in \mathbf{V}_h^{\boldsymbol{a}},\\
\left(\boldsymbol{\nabla}\cdot\boldsymbol{\sigma}_{\alpha h,\mathrm{disc}}^{k,i,\boldsymbol{a}}, q_h\right)_{\omega_h^{\boldsymbol{a}}} &= \left(\tilde{g}_{\alpha h}^{k,i,\boldsymbol{a}}, q_h\right)_{\omega_h^{\boldsymbol{a}}} && \forall q_h \in Q_h^{\boldsymbol{a}},
\end{aligned} \tag{36}$$

where the right-hand sides are defined by

$$\tilde{g}_{\alpha h}^{k,i,\boldsymbol{a}} := \left(f_\alpha - (-1)^\alpha\tilde{\lambda}_{h,\boldsymbol{a}}^{k,i} - r_{\alpha h}^{k,i}\right)\psi_{h,\boldsymbol{a}} - \mu_\alpha\boldsymbol{\nabla}u_{\alpha h}^{k,i}\cdot\boldsymbol{\nabla}\psi_{h,\boldsymbol{a}} \quad \forall\boldsymbol{a} \in \mathcal{V}_h, \tag{37}$$

recalling the notation (34)–(35). Then set

$$\boldsymbol{\sigma}_{1h,\mathrm{disc}}^{k,i} := \sum_{\boldsymbol{a}\in\mathcal{V}_h}\boldsymbol{\sigma}_{1h,\mathrm{disc}}^{k,i,\boldsymbol{a}} \quad \textit{and} \quad \boldsymbol{\sigma}_{2h,\mathrm{disc}}^{k,i} := \sum_{\boldsymbol{a}\in\mathcal{V}_h}\boldsymbol{\sigma}_{2h,\mathrm{disc}}^{k,i,\boldsymbol{a}}. \tag{38}$$

Consider an interior vertex $\boldsymbol{a} \in \mathcal{V}_h^{\mathrm{i}}$ and take the $\mathbb{P}_1$ hat basis function $\psi_{h,\boldsymbol{a}}$ in (33). This shows that $(\tilde{g}_{1h}^{k,i,\boldsymbol{a}}, 1)_{\omega_h^{\boldsymbol{a}}} = 0$, *i.e.*, the Neumann compatibility condition is satisfied for problems (36). Consequently, the second line of (36) holds true for all $q_h \in \mathbb{P}_p(\mathcal{T}_h|_{\omega_h^{\boldsymbol{a}}})$ (and not only on $Q_h^{\boldsymbol{a}}$). Since the functions $\psi_{h,\boldsymbol{a}}$ form a partition of unity as $\sum_{\boldsymbol{a}\in\mathcal{V}_h}\psi_{h,\boldsymbol{a}} = 1$ and since $r_{\alpha h}^{k,i}|_K$ and $\lambda_h^{k,i}|_K$ belong to $\mathbb{P}_p(K)$ for all $K \in \mathcal{T}_h$, we immediately have from [24, Lemma 3.5] that, for $\alpha \in \{1,2\}$,

$$\boldsymbol{\sigma}_{\alpha h,\mathrm{disc}}^{k,i} \in \mathbf{RT}_p(\Omega) \subset \mathbf{H}(\mathrm{div},\Omega) \ \text{ with } \ \boldsymbol{\nabla}\cdot\boldsymbol{\sigma}_{\alpha h,\mathrm{disc}}^{k,i} = \boldsymbol{\Pi}_{\mathbb{P}_p}(f_\alpha) - (-1)^\alpha\lambda_h^{k,i} - r_{\alpha h}^{k,i}. \tag{39}$$

Note that we use in particular $\sum_{\boldsymbol{a}\in\mathcal{V}_K}(\lambda_h^{k,i}\psi_{h,\boldsymbol{a}})|_K = \lambda_h^{k,i}|_K$ for $p \geq 2$ by the partition of unity by $\sum_{\boldsymbol{a}\in\mathcal{V}_h}\psi_{h,\boldsymbol{a}}|_K = 1|_K$, whereas $\sum_{\boldsymbol{a}\in\mathcal{V}_h}\tilde{\lambda}_{h,\boldsymbol{a}}^{k,i}\psi_{h,\boldsymbol{a}} = \sum_{\boldsymbol{a}\in\mathcal{V}_h}\lambda_h^{k,i}(\boldsymbol{a})\psi_{h,\boldsymbol{a}} = \lambda_h^{k,i}$ by the definition of the Lagrange basis for $p = 1$.

4.3 Algebraic error flux reconstruction via a multilevel approach

Given the piecewise polynomials $r_{\alpha h}^{k,i} \in \mathbb{P}_p(\mathcal{T}_h)$ from Section 4.1, we can immediately use the approach of [44, Concept 4.1] to obtain

$$\boldsymbol{\sigma}_{\alpha h,\mathrm{alg}}^{k,i} \in \mathbf{RT}_p(\Omega) \subset \mathbf{H}(\mathrm{div},\Omega) \ \text{ with } \ \boldsymbol{\nabla}\cdot\boldsymbol{\sigma}_{\alpha h,\mathrm{alg}}^{k,i} = r_{\alpha h}^{k,i}, \quad \forall\alpha \in \{1,2\}. \tag{40}$$

This construction requires a hierarchy of nested meshes of Ω and corresponds to one step of V-cycle multigrid. Setting $\boldsymbol{\sigma}_{\alpha h}^{k,i} := \boldsymbol{\sigma}_{\alpha h,\mathrm{alg}}^{k,i} + \boldsymbol{\sigma}_{\alpha h,\mathrm{disc}}^{k,i}$, (39) with (40) yields (32). [—]

5 A posteriori error estimates

We derive in this section an a posteriori estimate on the error between the exact solution $\boldsymbol{u}$ and the approximate solution $\boldsymbol{u}_h^{k,i}$ valid at each linearization iteration $k \geq 1$ and each algebraic iteration $i \geq 1$ of any inexact semismooth Newton method of Section 3. The main difficulty lies in the treatment of the constraints: the conditions $u_{1h}^{k,i} - u_{2h}^{k,i} \geq 0$ and $\lambda_h^{k,i} \geq 0$ do not necessarily hold before the convergence of both solvers for $p = 1$, and not even at convergence for $p \geq 2$. To treat the possible case $\lambda_h^{k,i} < 0$, let

$$\lambda_h^{k,i} = \lambda_h^{k,i,\mathrm{pos}} + \lambda_h^{k,i,\mathrm{neg}}, \quad \lambda_h^{k,i,\mathrm{pos}} := \max\{\lambda_h^{k,i}, 0\}, \quad \lambda_h^{k,i,\mathrm{neg}} := \min\{\lambda_h^{k,i}, 0\}$$

be its positive and negative parts. Note that $\lambda_h^{k,i,\mathrm{pos}} \in \Lambda$ but in general $\lambda_h^{k,i,\mathrm{pos}}, \lambda_h^{k,i,\mathrm{neg}} \notin X_h^p$ when $p \geq 2$.

5.1 A guaranteed a posteriori error estimate for the displacements

For local elementwise estimators $\eta_{\cdot,K}^{k,i}$, $K \in \mathcal{T}_h$, let their global counterparts be $\eta_{\cdot}^{k,i} := \big\{ \sum_{K\in\mathcal{T}_h} (\eta_{\cdot,K}^{k,i})^2 \big\}^{\frac12}$. Let $C_{\Omega,\mu} := h_\Omega C_{\mathrm{PF}} \big(\frac{1}{\mu_1} + \frac{1}{\mu_2}\big)^{\frac12}$. The first main result of this article is:

Theorem 5.1 (Guaranteed a posteriori estimate for the displacements)**.** *Let $\boldsymbol{u} = (u_1, u_2) \in \boldsymbol{\mathcal{K}}_g$ be the solution of the continuous reduced problem (13). Let $\boldsymbol{u}_h^{k,i} = (u_{1h}^{k,i}, u_{2h}^{k,i}) \in X_{gh}^p \times X_{0h}^p$ and $\lambda_h^{k,i} \in X_h^p$ be the approximation given by (31) for any $p \geq 1$, any linearization step $k \geq 1$, and any algebraic solver step $i \geq 1$. Let $\boldsymbol{\sigma}_{1h}^{k,i}$ and $\boldsymbol{\sigma}_{2h}^{k,i}$ be the equilibrated flux reconstructions of Section 4. Let finally $\tilde{\boldsymbol{s}}_h^{k,i} \in \boldsymbol{\mathcal{K}}_g$ be arbitrary. For $\alpha \in \{1,2\}$, define the estimators*

$$\eta_{\mathrm{F},K,\alpha}^{k,i} := \left\| \mu_\alpha^{\frac12} \boldsymbol{\nabla} u_{\alpha h}^{k,i} + \mu_\alpha^{-\frac12} \boldsymbol{\sigma}_{\alpha h}^{k,i} \right\|_K, \quad \eta_{\mathrm{osc},K,\alpha} := \frac{h_K}{\pi} \mu_\alpha^{-\frac12} \left\| f_\alpha - \boldsymbol{\Pi}_{\mathbb{P}_p}(f_\alpha) \right\|_K,$$

$$\eta_{\mathrm{C},K}^{k,i,\mathrm{pos}} := 2 \left(\lambda_h^{k,i,\mathrm{pos}}, u_{1h}^{k,i} - u_{2h}^{k,i} \right)_K, \quad \eta_1^{k,i} := \left(\sum_{K\in\mathcal{T}_h} \sum_{\alpha=1}^{2} \left(\eta_{\mathrm{F},K,\alpha}^{k,i} + \eta_{\mathrm{osc},K,\alpha} \right)^2 \right)^{\frac12},$$

$$\eta_{\mathrm{nonc},1,K}^{k,i} := \left| \left| \left| \tilde{\boldsymbol{s}}_h^{k,i} - \boldsymbol{u}_h^{k,i} \right| \right| \right|_K, \qquad \eta_{\mathrm{nonc},2,K}^{k,i} := C_{\Omega,\mu} \left\| \lambda_h^{k,i,\mathrm{neg}} \right\|_K,$$

$$\eta_{\mathrm{nonc},3,K}^{k,i} := 2 C_{\Omega,\mu} \left\| \lambda_h^{k,i,\mathrm{pos}} \right\| \left| \left| \left| \tilde{\boldsymbol{s}}_h^{k,i} - \boldsymbol{u}_h^{k,i} \right| \right| \right|_K.$$

Then, the following a posteriori error estimate holds:

$$\left| \left| \left| \boldsymbol{u} - \boldsymbol{u}_h^{k,i} \right| \right| \right| \leq \eta^{k,i} := \left\{ \left(\eta_1^{k,i} + \eta_{\mathrm{nonc},1}^{k,i} + \eta_{\mathrm{nonc},2}^{k,i} \right)^2 + \eta_{\mathrm{nonc},3}^{k,i} + \sum_{K\in\mathcal{T}_h} \eta_{\mathrm{C},K}^{k,i,\mathrm{pos}} \right\}^{\frac12}. \tag{41}$$

Remark 5.2. *The estimators of Theorem 5.1 reflect various violations of physical properties of the approximate solution $(\boldsymbol{u}_h^{k,i}, \lambda_h^{k,i})$: $\eta_{\mathrm{F},K,\alpha}^{k,i}$ and $\eta_{\mathrm{osc},K,\alpha}$ represent the nonconformity of the flux,* i.e.*, the fact that $-\mu_\alpha \boldsymbol{\nabla} u_{\alpha h}^{k,i} \notin \mathbf{H}(\mathrm{div},\Omega)$; $\eta_{\mathrm{C},K}^{k,i,\mathrm{pos}}$ reflects inconsistencies in the contact conditions at the discrete level,* i.e.*, the fact that $(u_{1h}^{k,i} - u_{2h}^{k,i})\lambda_h^{k,i} \neq 0$ everywhere in Ω; $\eta_{\mathrm{nonc},1,K}^{k,i}$, $\eta_{\mathrm{nonc},2,K}^{k,i}$, and $\eta_{\mathrm{nonc},3,K}^{k,i}$ stem from the possible departure of the discrete solution $\boldsymbol{u}_h^{k,i}$ from the convex set $\boldsymbol{\mathcal{K}}_g$ and the possible negativity of the discrete Lagrange multiplier $\lambda_h^{k,i}$. [—]*

Remark 5.3. *In [6], an a posteriori estimate between the exact solution $\boldsymbol{u}$ and the finite element approximation $\boldsymbol{u}_h$ given by (15) for $p = 1$, not taking into account nonlinear and linear solvers, was derived. Estimate (41) is its consistent extension to the present setting. [—]*

Proof of Theorem 5.1. First, as $\boldsymbol{u}_h^{k,i}$ does not belong to $\boldsymbol{\mathcal{K}}_g$ in general, we define the a-orthogonal projection $\boldsymbol{s} \in \boldsymbol{\mathcal{K}}_g$ of $\boldsymbol{u}_h^{k,i}$ to the nonempty closed convex set $\boldsymbol{\mathcal{K}}_g$ by

$$a(\boldsymbol{s}, \boldsymbol{v} - \boldsymbol{s}) \geq a(\boldsymbol{u}_h^{k,i}, \boldsymbol{v} - \boldsymbol{s}) \quad \forall \boldsymbol{v} \in \boldsymbol{\mathcal{K}}_g, \tag{42}$$

where we recall that the bilinear symmetric form a was defined in (10). Problem (42) is well-posed thanks to the Lions–Stampacchia theorem [41], because a defines a scalar product on $[H_0^1(\Omega)]^2$. Developing the square, the projection $\boldsymbol{s}$ satisfies for each $\boldsymbol{v} \in \boldsymbol{\mathcal{K}}_g$

$$\left|\!\left|\!\left| \boldsymbol{v} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right|^2 = |||\boldsymbol{v} - \boldsymbol{s}|||^2 + 2a(\boldsymbol{v} - \boldsymbol{s}, \boldsymbol{s} - \boldsymbol{u}_h^{k,i}) + \left|\!\left|\!\left| \boldsymbol{s} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right|^2 . \tag{43}$$

Since $a(\boldsymbol{v} - \boldsymbol{s}, \boldsymbol{s} - \boldsymbol{u}_h^{k,i}) \geq 0$ from (42), taking successively in (43) $\boldsymbol{v} = \boldsymbol{u}$ and $\boldsymbol{v} = \tilde{\boldsymbol{s}}_h^{k,i}$ for any $\tilde{\boldsymbol{s}}_h^{k,i} \in \boldsymbol{\mathcal{K}}_g$, we obtain

$$|||\boldsymbol{u} - \boldsymbol{s}||| \leq \left|\!\left|\!\left| \boldsymbol{u} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right| , \tag{44}$$

$$\left|\!\left|\!\left| \boldsymbol{s} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right| \leq \left|\!\left|\!\left| \tilde{\boldsymbol{s}}_h^{k,i} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right| = \eta_{\mathrm{nonc},1}^{k,i}. \tag{45}$$

Second, the energy norm of the error is decomposed as

$$\left|\!\left|\!\left| \boldsymbol{u} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right|^2 = a(\boldsymbol{u} - \boldsymbol{u}_h^{k,i}, \boldsymbol{u} - \boldsymbol{u}_h^{k,i}) = a(\boldsymbol{u} - \boldsymbol{u}_h^{k,i}, \boldsymbol{u} - \boldsymbol{s}) + a(\boldsymbol{u} - \boldsymbol{u}_h^{k,i}, \boldsymbol{s} - \boldsymbol{u}_h^{k,i}). \tag{46}$$

We estimate both terms in (46) separately. The second one is bounded by the Cauchy–Schwarz inequality and (45),

$$a(\boldsymbol{u} - \boldsymbol{u}_h^{k,i}, \boldsymbol{s} - \boldsymbol{u}_h^{k,i}) \leq \left|\!\left|\!\left| \boldsymbol{u} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right| \left|\!\left|\!\left| \boldsymbol{s} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right| \leq \left|\!\left|\!\left| \boldsymbol{u} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right| \eta_{\mathrm{nonc},1}^{k,i}. \tag{47}$$

The rest of the proof is dedicated to bounding the first one.

The reduced problem (13) for $\boldsymbol{v} = \boldsymbol{s} \in \boldsymbol{\mathcal{K}}_g$ yields

$$a(\boldsymbol{u}, \boldsymbol{u} - \boldsymbol{s}) \leq l(\boldsymbol{u} - \boldsymbol{s}). \tag{48}$$

Setting $\boldsymbol{w} = \boldsymbol{u} - \boldsymbol{s}$, we estimate the first term in (46) using (48) and adding and substracting $b(\boldsymbol{w}, \lambda_h^{k,i})$ and employing the definitions of b and l of (10)

$$\begin{aligned} a(\boldsymbol{u} - \boldsymbol{u}_h^{k,i}, \boldsymbol{w}) &\leq l(\boldsymbol{w}) + b(\boldsymbol{w}, \lambda_h^{k,i}) - a(\boldsymbol{u}_h^{k,i}, \boldsymbol{w}) - b(\boldsymbol{w}, \lambda_h^{k,i}), \\ &= \sum_{\alpha=1}^{2} \left(f_\alpha - (-1)^\alpha \lambda_h^{k,i}, w_\alpha \right) - \sum_{\alpha=1}^{2} \left(\mu_\alpha \boldsymbol{\nabla} u_{\alpha h}^{k,i}, \boldsymbol{\nabla} w_\alpha \right) - b(\boldsymbol{w}, \lambda_h^{k,i}). \end{aligned} \tag{49}$$

Besides, as $\boldsymbol{\sigma}_{\alpha h}^{k,i} \in \mathbf{H}(\mathrm{div}, \Omega)$ and since $w_\alpha \in H_0^1(\Omega)$, the Green formula gives

$$\left(\boldsymbol{\nabla} \cdot \boldsymbol{\sigma}_{\alpha h}^{k,i}, w_\alpha \right) = - \left(\boldsymbol{\sigma}_{\alpha h}^{k,i}, \boldsymbol{\nabla} w_\alpha \right) \quad \forall \alpha \in \{1, 2\} . \tag{50}$$

Then, using (50) in (49), one has

$$\begin{aligned} a(\boldsymbol{u} - \boldsymbol{u}_h^{k,i}, \boldsymbol{w}) \leq \sum_{\alpha=1}^{2} \sum_{K \in \mathcal{T}_h} \Big\{ &\left(f_\alpha - (-1)^\alpha \lambda_h^{k,i} - \boldsymbol{\nabla} \cdot \boldsymbol{\sigma}_{\alpha h}^{k,i}, w_\alpha \right)_K \\ &- \left(\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} u_{\alpha h}^{k,i} + \mu_\alpha^{-\frac{1}{2}} \boldsymbol{\sigma}_{\alpha h}^{k,i}, \mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} w_\alpha \right)_K \Big\} - b(\boldsymbol{w}, \lambda_h^{k,i}). \end{aligned} \tag{51}$$

It remains to bound each of the three terms in (51).

Using the divergence property (32), the Cauchy–Schwarz and Poincaré–Wirtinger (7b) inequalities, since $w_\alpha \in H^1(K)$, and denoting by $\overline{w}_{\alpha,K}$ the mean of w_α over K, for $\alpha = 1, 2$,

$$\left(f_\alpha - \boldsymbol{\nabla} \cdot \boldsymbol{\sigma}_{\alpha h}^{k,i} - (-1)^\alpha \lambda_h^{k,i}, w_\alpha \right)_K = \left(f_\alpha - \boldsymbol{\Pi}_{\mathbb{P}_p}(f_\alpha), w_\alpha - \overline{w}_{\alpha,K} \right)_K , \leq \eta_{\mathrm{osc},K,\alpha} \left\| \mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} w_\alpha \right\|_K . \tag{52}$$

Furthermore, by the Cauchy–Schwarz inequality

$$- \left(\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} u_{\alpha h}^{k,i} + \mu_\alpha^{-\frac{1}{2}} \boldsymbol{\sigma}_{\alpha h}^{k,i}, \mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} w_\alpha \right)_K \leq \eta_{\mathrm{F},K,\alpha}^{k,i} \left\| \mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} w_\alpha \right\|_K . \tag{53}$$

Next, as $\boldsymbol{u} \in \boldsymbol{\mathcal{K}}_g$, $\boldsymbol{w} = \boldsymbol{u} - \boldsymbol{s}$, and $-b(\boldsymbol{u}, \lambda_h^{k,i,\mathrm{pos}}) \le 0$, we have

$$\begin{aligned} -b(\boldsymbol{w}, \lambda_h^{k,i}) &\le -b(\boldsymbol{w}, \lambda_h^{k,i,\mathrm{neg}}) + b(\boldsymbol{s} - \boldsymbol{u}_h^{k,i}, \lambda_h^{k,i,\mathrm{pos}}) + b(\boldsymbol{u}_h^{k,i}, \lambda_h^{k,i,\mathrm{pos}}) \\ &= -\left(\lambda_h^{k,i,\mathrm{neg}}, w_1 - w_2\right) + \left(\lambda_h^{k,i,\mathrm{pos}}, (s_1 - u_{1h}^{k,i}) - (s_2 - u_{2h}^{k,i})\right) \\ &\quad + \frac{1}{2} \sum_{K \in \mathcal{T}_h} 2 \left(\lambda_h^{k,i,\mathrm{pos}}, u_{1h}^{k,i} - u_{2h}^{k,i}\right)_K . \end{aligned}$$

Using (8), we see

$$\|\boldsymbol{\nabla}(w_1 - w_2)\| \le \sum_{\alpha=1}^{2} \mu_\alpha^{-\frac{1}{2}} \left\| \mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} w_\alpha \right\| \le \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right)^{\frac{1}{2}} |||\boldsymbol{w}|||.$$

Thus, the Cauchy–Schwarz and Poincaré–Friedrichs (7a) inequalities, noting that both w_α and $(s_\alpha - u_{\alpha h}^{k,i})$ belong to $H_0^1(\Omega)$, and also employing (45), we obain

$$-b(\boldsymbol{w}, \lambda_h^{k,i}) \le \eta_{\mathrm{nonc},2}^{k,i} |||\boldsymbol{w}||| + \frac{1}{2} \eta_{\mathrm{nonc},3}^{k,i} + \frac{1}{2} \sum_{K \in \mathcal{T}_h} \eta_{\mathrm{C},K}^{k,i,\mathrm{pos}}. \tag{54}$$

Therefore, combining (46), (47), (51), (52), (53), (54), and (44), we get

$$\left|\left|\left| \boldsymbol{u} - \boldsymbol{u}_h^{k,i} \right|\right|\right|^2 \le \left(\eta_{\mathrm{nonc},1}^{k,i} + \eta_1^{k,i} + \eta_{\mathrm{nonc},2}^{k,i} \right) \left|\left|\left| \boldsymbol{u} - \boldsymbol{u}_h^{k,i} \right|\right|\right| + \frac{1}{2} \eta_{\mathrm{nonc},3}^{k,i} + \frac{1}{2} \sum_{K \in \mathcal{T}_h} \eta_{\mathrm{C},K}^{k,i,\mathrm{pos}}.$$

To conclude, the inequality $AB \le \frac{1}{2}(A^2 + B^2)$ gives the result (41). □

5.2 Construction of $\tilde{\boldsymbol{s}}_h^{k,i}$

As for the choice of $\tilde{\boldsymbol{s}}_h^{k,i} \in \boldsymbol{\mathcal{K}}_g$ in Theorem 5.1, a possibility is to proceed as follows. In addition to $\boldsymbol{\mathcal{K}}_{gh}^p$ of (14), for a polynomial degree $p' \ge p$, introduce the convex set

$$\widetilde{\boldsymbol{\mathcal{K}}}_{gh}^{p'} := \left\{ (v_{1h}, v_{2h}) \in X_{gh}^{p'} \times X_{0h}^{p'},\ v_{1h} - v_{2h} \ge 0 \right\} \subset \boldsymbol{\mathcal{K}}_g; \tag{55}$$

note that for $p' = 1$, $\widetilde{\boldsymbol{\mathcal{K}}}_{gh}^1 = \boldsymbol{\mathcal{K}}_{gh}^1$ but $\widetilde{\boldsymbol{\mathcal{K}}}_{gh}^p \subsetneq \boldsymbol{\mathcal{K}}_{gh}^p$ for $p \ge 2$. From $\boldsymbol{u}_h^{k,i} = \left(u_{1h}^{k,i}, u_{2h}^{k,i}\right) \in X_{gh}^p \times X_{0h}^p \not\subset \boldsymbol{\mathcal{K}}_{gh}^p$, we then first construct $\boldsymbol{s}_h^{k,i} \in \boldsymbol{\mathcal{K}}_{gh}^p$ such that, $\forall \boldsymbol{x}_l \in \mathcal{V}^{p,\mathrm{i}}$,

$$\boldsymbol{s}_h^{k,i}(\boldsymbol{x}_l) := \begin{cases} \boldsymbol{u}_h^{k,i}(\boldsymbol{x}_l) = \left(u_{1h}^{k,i}(\boldsymbol{x}_l), u_{2h}^{k,i}(\boldsymbol{x}_l)\right) & \text{if } u_{1h}^{k,i}(\boldsymbol{x}_l) \ge u_{2h}^{k,i}(\boldsymbol{x}_l), \\ \left(\frac{1}{2}\left(u_{1h}^{k,i}(\boldsymbol{x}_l) + u_{2h}^{k,i}(\boldsymbol{x}_l)\right), \frac{1}{2}\left(u_{1h}^{k,i}(\boldsymbol{x}_l) + u_{2h}^{k,i}(\boldsymbol{x}_l)\right) \right) & \text{if } u_{1h}^{k,i}(\boldsymbol{x}_l) < u_{2h}^{k,i}(\boldsymbol{x}_l). \end{cases} \tag{56}$$

When $p = 1$, we can take $\tilde{\boldsymbol{s}}_h^{k,i} := \boldsymbol{s}_h^{k,i}$, leading to the requested $\tilde{\boldsymbol{s}}_h^{k,i} \in \boldsymbol{\mathcal{K}}_{gh}^1 = \widetilde{\boldsymbol{\mathcal{K}}}_{gh}^1 \subset \boldsymbol{\mathcal{K}}_g$.

When $p \ge 2$, it may happen that even if the first components of $\boldsymbol{s}_h^{k,i}$ are greater or equal to the second components of $\boldsymbol{s}_h^{k,i}$ in the Lagrange nodes, $s_{1h}^{k,i} \not\ge s_{2h}^{k,i}$ everywhere, so that $\boldsymbol{s}_h^{k,i} \notin \widetilde{\boldsymbol{\mathcal{K}}}_{gh}^p$. We then employ the following procedure:

1. Go through all edges e of the mesh $\mathcal{T}_h$ lying in the interior of the domain Ω, $e \in \mathcal{E}_h^i$.
   (a) Consider $s_e := \left(s_{1h}^{k,i} - s_{2h}^{k,i}\right)|_e$. This is a p-degree polynomial on the one-dimensional segment e. If $s_e \ge 0$, set $c_e := 0$. Otherwise s_e takes negative values inside e but is non-negative at the two vertices of e by virtue of (56) .
   (b) Consider the edge bubble function ψ_e: this a non-negative piecewise second-order polynomial defined over ω_e, the subdomain formed by the two triangles that share the edge e, continuous over e, zero on $\partial\omega_e$, and with $\|\psi_e\|_{\infty,\omega_e} = 1$.
   (c) Let c_e be the smallest positive constant such that $(s_e + c_e \psi_e|_e) \ge 0$ on the edge e.
2. Go through all elements K of the mesh $\mathcal{T}_h$.

(a) Consider $s_K := \big(s_{1h}^{k,i} - s_{2h}^{k,i}\big)|_K + \big(\sum_{e\in\mathcal{E}_h^i} c_e\psi_e\big)|_K$. This is a p-degree polynomial on the two-dimensional triangle K. If $s_K \geq 0$, set $c_K := 0$. Otherwise s_K takes negative values inside K but is non-negative at the three edges of K.

(b) Consider the element bubble function ψ_K: this a non-negative third-order polynomial defined over K, zero on ∂K, and with $\|\psi_K\|_{\infty,K} = 1$.

(c) Let c_K be the smallest positive constant such that $s_K + c_K\psi_K \geq 0$ on K.

3. Define $\tilde{\boldsymbol{s}}_h^{k,i}$ by

$$\begin{aligned}\tilde{s}_{1h}^{k,i} &:= s_{1h}^{k,i} + \frac{1}{2}\sum_{e\in\mathcal{E}_h^i} c_e\psi_e + \frac{1}{2}\sum_{K\in\mathcal{T}_h} c_K\psi_K,\\ \tilde{s}_{2h}^{k,i} &:= s_{2h}^{k,i} - \frac{1}{2}\sum_{e\in\mathcal{E}_h^i} c_e\psi_e - \frac{1}{2}\sum_{K\in\mathcal{T}_h} c_K\psi_K.\end{aligned} \tag{57}$$

We easily see from the above that $\tilde{s}_{1h}^{k,i} \geq \tilde{s}_{2h}^{k,i}$, so that $\tilde{\boldsymbol{s}}_h^{k,i} \in \widetilde{\boldsymbol{\mathcal{K}}}_{gh}^{\max\{p,3\}} \subset \boldsymbol{\mathcal{K}}_g$.

5.3 A guaranteed a posteriori error estimate for the actions

Concerning $\lambda_h^{k,i}$, the following estimate holds (recall the definition given in (9)):

Theorem 5.4 (Guaranteed a posteriori estimate for the actions)**.** *Assume the hypotheses and notations of Theorem 5.1 and let $\lambda \in \Lambda$ be the solution of problem (11). Then*

$$\left|\!\left|\!\left|\lambda - \lambda_h^{k,i}\right|\!\right|\!\right|_{H_*^{-1}(\Omega)} \leq \eta^{k,i} + \eta_1^{k,i}. \tag{58}$$

Proof. The proof follows the one in [6, Corollary 3.5]. We only give the essential elements. Let $\mu_{\rm m} := \max(\mu_1,\mu_2)$. Employing (9) and extending appropriately b,

$$\left|\!\left|\!\left|\lambda - \lambda_h^{k,i}\right|\!\right|\!\right|_{H_*^{-1}(\Omega)} = \sup_{\substack{\psi\in H_0^1(\Omega)\\ \mu_{\rm m}\|\boldsymbol{\nabla}\psi\|^2=1}} \langle\lambda_h^{k,i} - \lambda, \psi\rangle = \sup_{\substack{\boldsymbol{\phi}\in[H_0^1(\Omega)]^2\\ \mu_{\rm m}\sum_{\alpha=1}^2\|\boldsymbol{\nabla}\phi_\alpha\|^2=1}} b(\boldsymbol{\phi}, \lambda_h^{k,i} - \lambda).$$

Fix $\boldsymbol{\phi} \in [H_0^1(\Omega)]^2$ such that $\mu_{\rm m}\sum_{\alpha=1}^2 \|\boldsymbol{\nabla}\phi_\alpha\|^2 = 1$. Invoking (11), we have

$$-b(\boldsymbol{\phi}, \lambda - \lambda_h^{k,i}) = l(\boldsymbol{\phi}) + b(\boldsymbol{\phi}, \lambda_h^{k,i}) - a(\boldsymbol{u}_h^{k,i}, \boldsymbol{\phi}) - a(\boldsymbol{u} - \boldsymbol{u}_h^{k,i}, \boldsymbol{\phi}).$$

The last term is estimated as $-a(\boldsymbol{u} - \boldsymbol{u}_h^{k,i}, \boldsymbol{\phi}) \leq \left|\!\left|\!\left|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\right|\!\right|\!\right|$, since $|||\boldsymbol{\phi}||| \leq 1$. The first three terms are identical to the first three terms of (49) but with $\boldsymbol{\phi} \in [H_0^1(\Omega)]^2$ instead of $\boldsymbol{w}$. Thus, using the estimates (52) and (53), one gets

$$-b(\boldsymbol{\phi}, \lambda - \lambda_h^{k,i}) \leq \eta_1^{k,i} + \left|\!\left|\!\left|\boldsymbol{u} - \boldsymbol{u}_h^{k,i}\right|\!\right|\!\right|,$$

which combined with (41) gives the result. □

Remark 5.5. *At convergence, for $\mathbb{P}_1$ finite elements, estimate (58) reduces to (3.30) in [6] with a slightly sharper treatment of the oscillation in f_α.*

5.4 Distinguishing the different error components

We now distinguish the different error components in the estimators from Theorem 5.1, by identifying the discretization estimator $\eta_{\rm disc}^{k,i}$, the semismooth linearization estimator $\eta_{\rm lin}^{k,i}$, and the linear algebra estimator $\eta_{\rm alg}^{k,i}$, such that $\eta_{\rm alg}^{k,i} \to 0$ when $i \to \infty$, $\eta_{\rm lin}^{k,i} \to 0$ and $\eta_{\rm alg}^{k,i} \to 0$ when $k \to \infty$ and $i \to \infty$, and all $\eta_{\rm disc}^{k,i}, \eta_{\rm lin}^{k,i}, \eta_{\rm alg}^{k,i} \to 0$ when $h \to 0$, $k \to \infty$, $i \to \infty$, supposing that $\boldsymbol{u}_h \to \boldsymbol{u} \in \boldsymbol{\mathcal{K}}_g$ and $\lambda_h \to \lambda \in \Lambda$, which was proven for $p = 1$ in [5].

When $p = 1$, the nonconformity estimators $\eta_{\mathrm{nonc},\beta}^{k,i}$, $1 \le \beta \le 3$, can be interpreted as estimators stemming from the semismooth linearization since they all tend to zero when $k \to \infty$ and $i \to \infty$; indeed, at convergence, $\tilde{\boldsymbol{s}}_h^{k,i} = \boldsymbol{s}_h^{k,i} = \boldsymbol{u}_h$ from (56) and (22), whereas $\lambda_h^{k,i,\mathrm{neg}} = \lambda_h^{\mathrm{neg}} = 0$ from (23). The estimates $\eta_{C,K}^{k,i,\mathrm{pos}}$ are attributed to discretization, as they only vanish when $h \to 0$.

When $p \ge 2$, the triangle inequality gives

$$\left|\!\left|\!\left| \tilde{\boldsymbol{s}}_h^{k,i} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right| \le \underbrace{\left|\!\left|\!\left| \tilde{\boldsymbol{s}}_h^{k,i} - \boldsymbol{s}_h^{k,i} \right|\!\right|\!\right|}_{\text{discretization}} + \underbrace{\left|\!\left|\!\left| \boldsymbol{s}_h^{k,i} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right|}_{\text{linearization}}; \tag{59}$$

here, from constructions (56)–(57), the first term vanishes for $h \to 0$, $k \to \infty$, $i \to \infty$, whereas the other one for $k \to \infty$ and $i \to \infty$. Thanks to (21) and (16b), we can decompose

$$\left(\lambda_h^{k,i,\mathrm{pos}}, u_{1h}^{k,i} - u_{2h}^{k,i}\right) = \underbrace{\left(\lambda_h^{k,i,\mathrm{pos}} - \lambda_h^{k,i}, u_{1h}^{k,i} - u_{2h}^{k,i}\right)}_{\text{discretization}} + \underbrace{\left(\lambda_h^{k,i}, u_{1h}^{k,i} - u_{2h}^{k,i}\right)}_{\text{linearization}}. \tag{60}$$

Finally, using (25), we decompose $\lambda_h^{k,i} = \tilde{\lambda}_h^{k,i,\mathrm{pos}} + \tilde{\lambda}_h^{k,i,\mathrm{neg}}$ with

$$\tilde{\lambda}_h^{k,i,\mathrm{pos}} := \sum_{l=1}^{\mathcal{N}^{p,\mathrm{i}}} \max\left\{\left(\boldsymbol{X}_{3h}^{k,i}\right)_l, 0\right\} \Theta_{h,\boldsymbol{x}_l}, \qquad \tilde{\lambda}_h^{k,i,\mathrm{neg}} := \sum_{l=1}^{\mathcal{N}^{p,\mathrm{i}}} \min\left\{\left(\boldsymbol{X}_{3h}^{k,i}\right)_l, 0\right\} \Theta_{h,\boldsymbol{x}_l},$$

so that $\tilde{\lambda}_h^{k,i,\mathrm{pos}}$ and $-\tilde{\lambda}_h^{k,i,\mathrm{neg}} \in \Lambda_h^p \subset X_h^p$ (recall that $\lambda_h^{k,i,\mathrm{pos}}, \lambda_h^{k,i,\mathrm{neg}} \notin X_h^p$ in general), and use

$$\left|\lambda_h^{k,i,\mathrm{neg}}\right| \le \underbrace{\left|\lambda_h^{k,i,\mathrm{neg}} - \tilde{\lambda}_h^{k,i,\mathrm{neg}}\right|}_{\text{discretization}} + \underbrace{\left|\tilde{\lambda}_h^{k,i,\mathrm{neg}}\right|}_{\text{linearization}}, \tag{61}$$

which is consistent with $p = 1$ where $\lambda_h^{k,i,\mathrm{neg}} = \tilde{\lambda}_h^{k,i,\mathrm{neg}}$. Note that $\tilde{\lambda}_h^{k,i,\mathrm{neg}} \to 0$ when $k \to \infty$, $i \to \infty$, and $\lambda_h^{k,i,\mathrm{neg}}$ vanishes when $h \to 0$.

Corollary 5.6 (A posteriori estimate distinguishing the different error components)**.** *Assume the hypotheses and notations of Theorem 5.1 in the case $p = 1$. Define, for $\alpha \in \{1, 2\}$,*

$$\eta_{\mathrm{disc},K,\alpha}^{k,i} := \left\| \mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} u_{\alpha h}^{k,i} + \mu_\alpha^{-\frac{1}{2}} \boldsymbol{\sigma}_{\alpha h,\mathrm{disc}}^{k,i} \right\|_K, \quad \eta_{\mathrm{alg},K,\alpha}^{k,i} := \left\| \mu_\alpha^{-\frac{1}{2}} \boldsymbol{\sigma}_{\alpha h,\mathrm{alg}}^{k,i} \right\|_K, \tag{62a}$$

$$\eta_{\mathrm{alg}}^{k,i} := \left\{ \sum_{\alpha=1}^{2} \sum_{K \in \mathcal{T}_h} (\eta_{\mathrm{alg},K,\alpha}^{k,i})^2 \right\}^{\frac{1}{2}}, \tag{62b}$$

and if $p = 1$

$$\eta_{\mathrm{disc}}^{k,i} := \left\{ \sum_{\alpha=1}^{2} \sum_{K \in \mathcal{T}_h} \left(\eta_{\mathrm{disc},K,\alpha}^{k,i} + \eta_{\mathrm{osc},K,\alpha} \right)^2 \right\}^{\frac{1}{2}} + \left\{ \left| \sum_{K \in \mathcal{T}_h} \eta_{\mathrm{C},K}^{k,i,\mathrm{pos}} \right| \right\}^{\frac{1}{2}}, \tag{62c}$$

$$\eta_{\mathrm{lin}}^{k,i} := \eta_{\mathrm{nonc},1}^{k,i} + \eta_{\mathrm{nonc},2}^{k,i} + \left(\eta_{\mathrm{nonc},3}^{k,i} \right)^{\frac{1}{2}}, \tag{62d}$$

whereas if $p \ge 2$

$$\begin{aligned} \eta_{\mathrm{disc}}^{k,i} := {} & \left\{ \sum_{\alpha=1}^{2} \sum_{K \in \mathcal{T}_h} \left(\eta_{\mathrm{disc},K,\alpha}^{k,i} + \eta_{\mathrm{osc},K,\alpha} \right)^2 \right\}^{\frac{1}{2}} + \left\{ 2 \left| \left(\lambda_h^{k,i,\mathrm{pos}} - \lambda_h^{k,i}, u_{1h}^{k,i} - u_{2h}^{k,i} \right) \right| \right\}^{\frac{1}{2}} \\ & + \left|\!\left|\!\left| \tilde{\boldsymbol{s}}_h^{k,i} - \boldsymbol{s}_h^{k,i} \right|\!\right|\!\right| + C_{\Omega,\mu} \left| \lambda_h^{k,i,\mathrm{neg}} - \tilde{\lambda}_h^{k,i,\mathrm{neg}} \right| + \left(2 C_{\Omega,\mu} \left| \lambda_h^{k,i,\mathrm{pos}} \right| \right)^{\frac{1}{2}} \left|\!\left|\!\left| \tilde{\boldsymbol{s}}_h^{k,i} - \boldsymbol{s}_h^{k,i} \right|\!\right|\!\right|^{\frac{1}{2}}, \end{aligned} \tag{62e}$$

$$\begin{aligned} \eta_{\mathrm{lin}}^{k,i} := {} & \left|\!\left|\!\left| \boldsymbol{s}_h^{k,i} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right| + C_{\Omega,\mu} \left| \tilde{\lambda}_h^{k,i,\mathrm{neg}} \right| + \left(2 C_{\Omega,\mu} \left| \lambda_h^{k,i,\mathrm{pos}} \right| \right)^{\frac{1}{2}} \left|\!\left|\!\left| \boldsymbol{s}_h^{k,i} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right|^{\frac{1}{2}} \\ & + \left\{ 2 \left| \left(\lambda_h^{k,i}, u_{1h}^{k,i} - u_{2h}^{k,i} \right) \right| \right\}^{\frac{1}{2}}. \end{aligned} \tag{62f}$$

Then,

$$\left|\!\left|\!\left| \boldsymbol{u} - \boldsymbol{u}_h^{k,i} \right|\!\right|\!\right| \le \eta_{\mathrm{disc}}^{k,i} + \eta_{\mathrm{lin}}^{k,i} + \eta_{\mathrm{alg}}^{k,i}.$$

Proof. As for $(A,B) \in \mathbb{R}_+ \times \mathbb{R}_+$, $(A+B)^{\frac{1}{2}} \le A^{\frac{1}{2}} + B^{\frac{1}{2}}$, we have from (41)

$$\eta^{k,i} \le \eta_1^{k,i} + \eta_{\mathrm{nonc},1}^{k,i} + \eta_{\mathrm{nonc},2}^{k,i} + \left(\eta_{\mathrm{nonc},3}^{k,i}\right)^{\frac{1}{2}} + \left(\left|\sum_{K\in\mathcal{T}_h} \eta_{\mathrm{C},K}^{k,i,\mathrm{pos}}\right|\right)^{\frac{1}{2}}.$$

Next, the definition of $\eta_1^{k,i}$ combined with the triangle (Minkowski) inequality to separate the algebraic $\eta_{\mathrm{alg},K,\alpha}^{k,i}$ and the discretization $\eta_{\mathrm{disc},K,\alpha}^{k,i}$ estimators gives

$$\eta_1^{k,i} \le \left(\sum_{K\in\mathcal{T}_h}\sum_{\alpha=1}^{2}\left(\eta_{\mathrm{disc},K,\alpha}^{k,i} + \eta_{\mathrm{osc},K,\alpha}\right)^2\right)^{\frac{1}{2}} + \left(\sum_{K\in\mathcal{T}_h}\sum_{\alpha=1}^{2}\left(\eta_{\mathrm{alg},K,\alpha}^{k,i}\right)^2\right)^{\frac{1}{2}},$$

which finishes the proof for $p=1$. For $p \ge 2$, we need to additionally invoke (59)–(61). □

6 Adaptive inexact semismooth Newton method using a posteriori stopping criteria

We propose in this section an adaptive inexact semismooth Newton method. In the spirit of [23], it is designed to only perform the linearization and algebraic resolution with minimal necessary precision and thus to avoid unnecessary iterations. We rely on Corollary 5.6 that estimates the size of the different error components and design adaptive stopping criteria for both linearization and algebraic solvers. The results of this section are for simplicity presented for $p=1$; extension to $p \ge 2$ is merely technical.

6.1 A posteriori stopping criteria

Recall that we employ a semismooth Newton method for nonlinear problem (28), yielding on each step $k \ge 1$ linear system (29) that we solve inexactly in the sense of (31). Let γ_{lin} and γ_{alg} be two positive parameters typically of order 0.1, representing the desired relative sizes of the algebraic and linearization errors. We propose the following a posteriori stopping criteria, balancing the algebraic, linearization, and discretization estimators of Corollary 5.6:

$$\text{(a) } \eta_{\mathrm{alg}}^{k,i} \le \gamma_{\mathrm{alg}} \max\left\{\eta_{\mathrm{disc}}^{k,i}, \eta_{\mathrm{lin}}^{k,i}\right\}, \qquad \text{(b) } \eta_{\mathrm{lin}}^{k,i} \le \gamma_{\mathrm{lin}} \eta_{\mathrm{disc}}^{k,i}. \tag{63}$$

Remark 6.1. *When $p=1$, for all mesh elements $K \in \mathcal{T}_h$, let $\gamma_{\mathrm{lin},K}$, $\gamma_{\mathrm{alg},K}$ be two fixed parameters, typically of order* 0.1*, representing the desired local relative sizes of the linearization and algebraic errors components. Following [23, 35], one can aim at the balance of all error components in each mesh cell in place of* (63)*, while simultaneously guaranteeing the global criteria* (63)*. These local criteria read*

$$\eta_{\mathrm{alg},\omega_h^{\boldsymbol{a}},\alpha}^{k,i} \le \min_{K\subset\omega_h^{\boldsymbol{a}}}\left\{\gamma_{\mathrm{alg},K}\max\left\{\eta_{\mathrm{disc},K,\alpha}^{k,i}, \eta_{\mathrm{lin},K}^{k,i}\right\}\right\} \qquad \forall \alpha \in \{1,2\}, \tag{64a}$$

$$\eta_{\mathrm{lin},K}^{k,i} \le \min_{\alpha\in\{1,2\}}\left\{\gamma_{\mathrm{lin},K}\eta_{\mathrm{disc},K,\alpha}^{k,i}\right\}, \tag{64b}$$

where

$$\eta_{\mathrm{lin},K}^{k,i} := \left(1 + \left(2C_{\Omega,\mu}\left\|\lambda_h^{k,i,\mathrm{pos}}\right\|\right)^{\frac{1}{2}}\left|\!\left|\!\left|\tilde{\boldsymbol{s}}_h^{k,i} - \boldsymbol{u}_h^{k,i}\right|\!\right|\!\right|^{-\frac{1}{2}}\right)\eta_{\mathrm{nonc},1,K}^{k,i} + \eta_{\mathrm{nonc},2,K}^{k,i}, \tag{65a}$$

$$\eta_{\mathrm{alg},\omega_h^{\boldsymbol{a}},\alpha}^{k,i} := \left\{\sum_{K\subset\omega_h^{\boldsymbol{a}}}\left(\eta_{\mathrm{alg},K,\alpha}^{k,i}\right)^2\right\}^{\frac{1}{2}} = \left\|\mu_\alpha^{-\frac{1}{2}}\boldsymbol{\sigma}_{\alpha h,\mathrm{alg}}^{k,i}\right\|_{\omega_h^{\boldsymbol{a}}}. \tag{65b}$$

The (complicated) form of $\eta_{\mathrm{lin},K}^{k,i}$ ensures that local criteria (64) *imply the global criteria* (63)*, and stems from the different scalings of $\eta_{\mathrm{nonc},1,K}^{k,i}$ and $\eta_{\mathrm{nonc},2,K}^{k,i}$ with respect to $\eta_{\mathrm{nonc},3,K}^{k,i}$ in Theorem 5.1. In particular, local efficiency for $p=1$ will be proven below based on* (64)*.*

Remark 6.2. *When $p \geq 2$, for the sake of brevity, we only consider locally the algebraic error component and require the local stopping criterion*

$$\eta^{k,i}_{\mathrm{alg},\omega_h^{\boldsymbol{a}},\alpha} \leq \min_{K \subset \omega_h^{\boldsymbol{a}}} \left\{ \gamma_{\mathrm{alg},K} \eta^{k,i}_{\mathrm{disc},K,\alpha} \right\} \qquad \forall \alpha \in \{1,2\} \tag{66}$$

in place of (64)*, where* $\eta^{k,i}_{\mathrm{alg},\omega_h^{\boldsymbol{a}},\alpha}$ *is given by* (65b)*.*

6.2 Adaptive inexact semismooth Newton algorithm

The adaptive version of the inexact semismooth Newton algorithm of Section 3.3 that we propose is as follows:

Algorithm 1 Adaptive inexact semismooth Newton algorithm

0. Choose an initial vector $\boldsymbol{X}_h^0 \in \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}}}$ and set $k = 1$.
1. From $\boldsymbol{X}_h^{k-1}$ define $\mathbb{A}^{k-1} \in \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}},3\mathcal{N}^{p,\mathrm{i}}}$ and $\boldsymbol{B}^{k-1} \in \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}}}$ by (30).
2. Consider the linear system

$$\mathbb{A}^{k-1} \boldsymbol{X}_h^k = \boldsymbol{B}^{k-1}. \tag{67}$$

3. Set $\boldsymbol{X}_h^{k,0} := \boldsymbol{X}_h^{k-1}$ as initial guess for the iterative linear solver, set $i := 0$.
4*a*. Perform $\nu \geq 1$ steps of a chosen linear solver for (67), starting from $\boldsymbol{X}_h^{k,i}$. This yields on step $i+\nu$ an approximation $\boldsymbol{X}_h^{k,i+\nu}$ to $\boldsymbol{X}_h^k$ satisfying

$$\mathbb{A}^{k-1} \boldsymbol{X}_h^{k,i+\nu} = \boldsymbol{B}^{k-1} - \boldsymbol{R}_h^{k,i+\nu}.$$

4*b*. Compute the estimators of Corollary 5.6 and check the stopping criterion for the linear solver in the form (63)(a). Set $i := i + \nu$. If satisfied, set $\boldsymbol{X}_h^k := \boldsymbol{X}_h^{k,i}$. If not go back to **4*a***.
5. Check the stopping criterion for the nonlinear solver in the form (63)(b). If satisfied, return $\boldsymbol{X}_h := \boldsymbol{X}_h^k$. If not, set $k := k + 1$ and go back to **1**.

7 Efficiency

We prove in this section local efficiency of our a posteriori error estimators, proceeding following [10, 23, 24, 44]. [—] In the case $p = 1$, we rely on the local stopping criteria (64); in the generic case $p \geq 2$, we do not address the local efficiency in the presence of an inexact linearization solver and rely on (66). We assume in the sequel for simplicity that f_1 and f_2 are piecewise $\mathbb{P}_p$ polynomials. This obviously yields $\eta_{\mathrm{osc},K,\alpha} = 0$, $\forall \alpha \in \{1,2\}$. We do not treat here the "complementarity" estimators $\eta^{k,i,\mathrm{pos}}_{\mathrm{C},K}$ that are typically numerically very small. Their local efficiency could be proven, when $p = 1$, along the lines of [6, Proposition 3.9].

7.1 Continuous-level problems with hat functions on patches

For each vertex $\boldsymbol{a} \in \mathcal{V}_h$, define the spaces

$$\begin{aligned} H^1_*(\omega_h^{\boldsymbol{a}}) &:= \left\{ v \in H^1(\omega_h^{\boldsymbol{a}});\ (v,1)_{\omega_h^{\boldsymbol{a}}} = 0 \right\} && \boldsymbol{a} \in \mathcal{V}_h^{\mathrm{i}}, \\ H^1_*(\omega_h^{\boldsymbol{a}}) &:= \left\{ v \in H^1(\omega_h^{\boldsymbol{a}});\ v = 0 \text{ on } \partial\omega_h^{\boldsymbol{a}} \cap \partial\Omega \right\} && \boldsymbol{a} \in \mathcal{V}_h^{\mathrm{e}}. \end{aligned}$$

Then there is a constant $C_{\mathrm{cont,PF}} > 0$ only depending on the shape regularity of $\mathcal{T}_h$ such that

$$\|\boldsymbol{\nabla}(\psi_{h,\boldsymbol{a}} v)\|_{\omega_h^{\boldsymbol{a}}} \leq C_{\mathrm{cont,PF}} \|\boldsymbol{\nabla} v\|_{\omega_h^{\boldsymbol{a}}} \quad \forall v \in H^1_*(\omega_h^{\boldsymbol{a}}), \tag{68}$$

see Braess *et al.* [10] or Ern and Vohralík [24]. Then, we have:

Lemma 7.1. *Let* (u_1, u_2, λ) *be the solution of* (11) *and let* $(u_{1h}^{k,i}, u_{2h}^{k,i}, \lambda_h^{k,i})$ *be the approximation given by* (31)*, verifying in particular* (33)*. Let* $\boldsymbol{a} \in \mathcal{V}_h$*, and for* $\alpha \in \{1,2\}$*, let* $\zeta_{\alpha,\boldsymbol{a}} \in H^1_*(\omega_h^{\boldsymbol{a}})$ *be the solution of*

$$(\mu_\alpha \boldsymbol{\nabla}\zeta_{\alpha,\boldsymbol{a}}, \boldsymbol{\nabla} v)_{\omega_h^{\boldsymbol{a}}} = \left(-\mu_\alpha \psi_{h,\boldsymbol{a}} \boldsymbol{\nabla} u_{\alpha h}^{k,i}, \boldsymbol{\nabla} v \right)_{\omega_h^{\boldsymbol{a}}} + \left(\tilde{g}_{\alpha h}^{k,i,\boldsymbol{a}}, v \right)_{\omega_h^{\boldsymbol{a}}} \quad \forall v \in H^1_*(\omega_h^{\boldsymbol{a}}), \tag{69}$$

where $\tilde{g}_{\alpha h}^{k,i,\boldsymbol{a}}$ *is defined in* (37). *Let* $\mu_{\mathrm{m}} := \max(\mu_1, \mu_2)$. *Then, for* $\alpha \in \{1,2\}$,

$$\left\|\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \zeta_{\alpha,\boldsymbol{a}}\right\|_{\omega_h^{\boldsymbol{a}}} \leq C_{\mathrm{cont,PF}} \left(\left\|\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \left(u_\alpha - u_{\alpha h}^{k,i}\right)\right\|_{\omega_h^{\boldsymbol{a}}} + \mu_{\mathrm{m}}^{\frac{1}{2}} \mu_\alpha^{-\frac{1}{2}} \left|\!\left|\!\left|\lambda - \tilde{\lambda}_{h,\boldsymbol{a}}^{k,i}\right|\!\right|\!\right|_{H_*^{-1}(\omega_h^{\boldsymbol{a}})} + \left\|\mu_\alpha^{-\frac{1}{2}} \boldsymbol{\sigma}_{\alpha h,\mathrm{alg}}^{k,i}\right\|_{\omega_h^{\boldsymbol{a}}} \right). \tag{70}$$

Proof. Let $\alpha \in \{1,2\}$. There holds

$$\left\|\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \zeta_{\alpha,\boldsymbol{a}}\right\|_{\omega_h^{\boldsymbol{a}}} = \sup_{v \in H_*^1(\omega_h^{\boldsymbol{a}}), \left\|\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} v\right\|_{\omega_h^{\boldsymbol{a}}} = 1} \left(\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \zeta_{\alpha,\boldsymbol{a}}, \mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} v\right)_{\omega_h^{\boldsymbol{a}}}. \tag{71}$$

Consider $v \in H_*^1(\omega_h^{\boldsymbol{a}})$ with $\left\|\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} v\right\|_{\omega_h^{\boldsymbol{a}}} = 1$. As $\zeta_{\alpha,\boldsymbol{a}}$ is the solution of (69), using in (11) the definition (37) and considering the test functions $(\psi_{h,\boldsymbol{a}} v, 0)$ and $(0, \psi_{h,\boldsymbol{a}} v)$, that crucially belong to $\left(H_0^1(\omega_h^{\boldsymbol{a}})\right)^2 \subset \left(H_0^1(\Omega)\right)^2$ due to the multiplication by the hat functions $\psi_{h,\boldsymbol{a}}$,

$$\left(\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \zeta_{\alpha,\boldsymbol{a}}, \mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} v\right)_{\omega_h^{\boldsymbol{a}}} = \left(\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \left(u_\alpha - u_{\alpha h}^{k,i}\right), \mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \left(\psi_{h,\boldsymbol{a}} v\right)\right)_{\omega_h^{\boldsymbol{a}}} + \left((-1)^\alpha \left(\lambda - \tilde{\lambda}_{h,\boldsymbol{a}}^{k,i}\right) - r_{\alpha h}^{k,i}, \psi_{h,\boldsymbol{a}} v\right)_{\omega_h^{\boldsymbol{a}}}. \tag{72}$$

Moreover, as $\psi_{h,\boldsymbol{a}} v \in H_0^1(\omega_h^{\boldsymbol{a}})$, $\boldsymbol{\sigma}_{\alpha h,\mathrm{alg}}^{k,i} \in \mathbf{H}(\mathrm{div}, \omega_h^{\boldsymbol{a}})$, and $\boldsymbol{\nabla} \cdot \boldsymbol{\sigma}_{\alpha h,\mathrm{alg}}^{k,i} = r_{\alpha h}^{k,i}$ by (40), the Green formula and the Cauchy–Schwarz inequality give

$$\left|\left(r_{\alpha h}^{k,i}, \psi_{h,\boldsymbol{a}} v\right)_{\omega_h^{\boldsymbol{a}}}\right| = \left|-\left(\mu_\alpha^{-\frac{1}{2}} \boldsymbol{\sigma}_{\alpha h,\mathrm{alg}}^{k,i}, \mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \left(\psi_{h,\boldsymbol{a}} v\right)\right)_{\omega_h^{\boldsymbol{a}}}\right| \leq \left\|\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \left(\psi_{h,\boldsymbol{a}} v\right)\right\|_{\omega_h^{\boldsymbol{a}}} \left\|\mu_\alpha^{-\frac{1}{2}} \boldsymbol{\sigma}_{\alpha h,\mathrm{alg}}^{k,i}\right\|_{\omega_h^{\boldsymbol{a}}}. \tag{73}$$

Multiplying and dividing $\left(\lambda - \tilde{\lambda}_{h,\boldsymbol{a}}^{k,i}, \psi_{h,\boldsymbol{a}} v\right)_{\omega_h^{\boldsymbol{a}}}$ by $\left\|\mu_{\mathrm{m}}^{\frac{1}{2}} \boldsymbol{\nabla} \left(\psi_{h,\boldsymbol{a}} v\right)\right\|_{\omega_h^{\boldsymbol{a}}}$ and using that $\psi_{h,\boldsymbol{a}} v \in H_0^1(\omega_h^{\boldsymbol{a}})$, which allows us to employ definition (9), we get

$$\left|\left(\lambda - \tilde{\lambda}_{h,\boldsymbol{a}}^{k,i}, \psi_{h,\boldsymbol{a}} v\right)_{\omega_h^{\boldsymbol{a}}}\right| \leq \left|\!\left|\!\left|\lambda - \tilde{\lambda}_{h,\boldsymbol{a}}^{k,i}\right|\!\right|\!\right|_{H_*^{-1}(\omega_h^{\boldsymbol{a}})} \mu_{\mathrm{m}}^{\frac{1}{2}} \mu_\alpha^{-\frac{1}{2}} \left\|\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \left(\psi_{h,\boldsymbol{a}} v\right)\right\|_{\omega_h^{\boldsymbol{a}}}. \tag{74}$$

Finally, the Cauchy–Schwarz inequality leads to

$$\left(\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \left(u_\alpha - u_{\alpha h}^{k,i}\right), \mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \left(\psi_{h,\boldsymbol{a}} v\right)\right)_{\omega_h^{\boldsymbol{a}}} \leq \left\|\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \left(u_\alpha - u_{\alpha h}^{k,i}\right)\right\|_{\omega_h^{\boldsymbol{a}}} \left\|\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \left(\psi_{h,\boldsymbol{a}} v\right)\right\|_{\omega_h^{\boldsymbol{a}}}. \tag{75}$$

The result now follows by combining (73), (74), and (75) with (68) together with (71). □

7.2 Local efficiency of the estimators

Recall the definition of $\zeta_{\alpha,\boldsymbol{a}}$ from (69) in Lemma 7.1. Following [10, 24], there exists a constant $C_{\mathrm{st}} > 0$ only depending on the shape regularity of the mesh $\mathcal{T}_h$ such that the discretization flux reconstructions $\boldsymbol{\sigma}_{\alpha h,\mathrm{disc}}^{k,i,\boldsymbol{a}}$ of Definition 4.1 satisfy

$$\left\|\mu_\alpha^{\frac{1}{2}} \psi_{h,\boldsymbol{a}} \boldsymbol{\nabla} u_{\alpha h}^{k,i} + \mu_\alpha^{-\frac{1}{2}} \boldsymbol{\sigma}_{\alpha h,\mathrm{disc}}^{k,i,\boldsymbol{a}}\right\|_{\omega_h^{\boldsymbol{a}}} \leq C_{\mathrm{st}} \left\|\mu_\alpha^{\frac{1}{2}} \boldsymbol{\nabla} \zeta_{\alpha,\boldsymbol{a}}\right\|_{\omega_h^{\boldsymbol{a}}}. \tag{76}$$

Our second main result is:

Theorem 7.2 (Efficiency of the a posteriori estimate)**.** *Let the flux reconstructions* $\boldsymbol{\sigma}_{\alpha h,\mathrm{disc}}^{k,i}$ *be given by Definition 4.1 and let* $\boldsymbol{\sigma}_{\alpha h,\mathrm{alg}}^{k,i}$ *satisfy* (40). *Let the local stopping criteria* (64) *be satisfied for the estimators of Corollary 5.6 for* $p = 1$ *and* (66) *for* $p \geq 2$. *Let finally the algebraic parameters* $\gamma_{\mathrm{alg},K}$ *be such that*

$$\begin{aligned} \gamma_{\mathrm{alg},K} &\leq \frac{1}{6 C_{\mathrm{st}} C_{\mathrm{cont,PF}} \max\left\{1, \gamma_{\mathrm{lin},K}\right\}} \quad \text{if } p = 1, \\ \gamma_{\mathrm{alg},K} &\leq \frac{1}{6 C_{\mathrm{st}} C_{\mathrm{cont,PF}}} \quad \text{if} \quad p \geq 2. \end{aligned} \tag{77}$$

Setting

$$\delta_K := 2C_{\mathrm{st}}C_{\mathrm{cont,PF}}\left(1+\gamma_{\mathrm{lin},K}+\gamma_{\mathrm{alg},K}\max\left\{1,\gamma_{\mathrm{lin},K}\right\}\right) \quad \textit{if } p=1$$

and

$$\delta_K := 2C_{\mathrm{st}}C_{\mathrm{cont,PF}}\left(1+\gamma_{\mathrm{alg},K}\right) \quad \textit{if} \quad p\geq 2,$$

we have for $\alpha\in\{1,2\}$

$$\eta^{k,i}_{\mathrm{disc},K,\alpha}+\eta^{k,i}_{\mathrm{lin},K}+\eta^{k,i}_{\mathrm{alg},K,\alpha}\leq \delta_K\sum_{\boldsymbol{a}\in\mathcal{V}_K}\left(\left\|\mu_\alpha^{\frac{1}{2}}\boldsymbol{\nabla}\left(u_\alpha-u^{k,i}_{\alpha h}\right)\right\|_{\omega_h^{\boldsymbol{a}}}+\mu_{\mathrm{m}}^{\frac{1}{2}}\mu_\alpha^{-\frac{1}{2}}\left|\!\left|\!\left|\lambda-\tilde{\lambda}^{k,i}_{h,\boldsymbol{a}}\right|\!\right|\!\right|_{H_*^{-1}(\omega_h^{\boldsymbol{a}})}\right) \quad \textit{if } p=1,$$

$$\begin{aligned}\eta^{k,i}_{\mathrm{F},K,\alpha}&=\left\|\mu_\alpha^{\frac{1}{2}}\boldsymbol{\nabla}u^{k,i}_{\alpha h}+\mu_\alpha^{-\frac{1}{2}}\boldsymbol{\sigma}^{k,i}_{\alpha h}\right\|_K\\&\leq \eta^{k,i}_{\mathrm{disc},K,\alpha}+\eta^{k,i}_{\mathrm{alg},K,\alpha}\\&\leq \delta_K\sum_{\boldsymbol{a}\in\mathcal{V}_K}\left(\left\|\mu_\alpha^{\frac{1}{2}}\boldsymbol{\nabla}\left(u_\alpha-u^{k,i}_{\alpha h}\right)\right\|_{\omega_h^{\boldsymbol{a}}}+\mu_{\mathrm{m}}^{\frac{1}{2}}\mu_\alpha^{-\frac{1}{2}}\left|\!\left|\!\left|\lambda-\tilde{\lambda}^{k,i}_{h}\right|\!\right|\!\right|_{H_*^{-1}(\omega_h^{\boldsymbol{a}})}\right) \quad \textit{if} \quad p\geq 2.\end{aligned}$$

Proof. We first treat the case $p=1$. Let $\alpha\in\{1,2\}$. First, the local criteria (64a) and (64b) and the definition of δ_K yield

$$\eta^{k,i}_{\mathrm{disc},K,\alpha}+\eta^{k,i}_{\mathrm{lin},K}+\eta^{k,i}_{\mathrm{alg},K,\alpha}\leq\frac{\delta_K}{2C_{\mathrm{st}}C_{\mathrm{cont,PF}}}\eta^{k,i}_{\mathrm{disc},K,\alpha}. \tag{78}$$

Next, definition (62a) and (38) which implies $\boldsymbol{\sigma}^{k,i}_{\alpha h,\mathrm{disc}}|_K=\sum_{\boldsymbol{a}\in\mathcal{V}_K}\boldsymbol{\sigma}^{k,i,\boldsymbol{a}}_{\alpha h,\mathrm{disc}}|_K$ together with the partition of unity $\sum_{\boldsymbol{a}\in\mathcal{V}_K}\psi_{h,\boldsymbol{a}}|_K=1|_K$ imply

$$\eta^{k,i}_{\mathrm{disc},K,\alpha}\leq\sum_{\boldsymbol{a}\in\mathcal{V}_K}\left\|\mu_\alpha^{\frac{1}{2}}\psi_{h,\boldsymbol{a}}\boldsymbol{\nabla}u^{k,i}_{\alpha h}+\mu_\alpha^{-\frac{1}{2}}\boldsymbol{\sigma}^{k,i,\boldsymbol{a}}_{\alpha h,\mathrm{disc}}\right\|_{\omega_h^{\boldsymbol{a}}},$$

where we have also enlarged the domain of the integral. Thus, stability (76) and energy lower bound (70) lead to

$$\eta^{k,i}_{\mathrm{disc},K,\alpha}\leq C_{\mathrm{st}}C_{\mathrm{cont,PF}}\sum_{\boldsymbol{a}\in\mathcal{V}_K}\left(\left\|\mu_\alpha^{\frac{1}{2}}\boldsymbol{\nabla}\left(u_\alpha-u^{k,i}_{\alpha h}\right)\right\|_{\omega_h^{\boldsymbol{a}}}+\mu_{\mathrm{m}}^{\frac{1}{2}}\mu_\alpha^{-\frac{1}{2}}\left|\!\left|\!\left|\lambda-\tilde{\lambda}^{k,i}_{h,\boldsymbol{a}}\right|\!\right|\!\right|_{H_*^{-1}(\omega_h^{\boldsymbol{a}})}+\left\|\mu_\alpha^{-\frac{1}{2}}\boldsymbol{\sigma}^{k,i}_{\alpha h,\mathrm{alg}}\right\|_{\omega_h^{\boldsymbol{a}}}\right). \tag{79}$$

Using successively the local criteria (64), and since any triangle has three vertices,

$$\sum_{\boldsymbol{a}\in\mathcal{V}_K}\left\|\mu_\alpha^{-\frac{1}{2}}\boldsymbol{\sigma}^{k,i}_{\alpha h,\mathrm{alg}}\right\|_{\omega_h^{\boldsymbol{a}}}=\sum_{\boldsymbol{a}\in\mathcal{V}_K}\eta^{k,i}_{\mathrm{alg},\omega_h^{\boldsymbol{a}},\alpha}\leq 3\gamma_{\mathrm{alg},K}\max\left\{\eta^{k,i}_{\mathrm{disc},K,\alpha},\eta^{k,i}_{\mathrm{lin},K}\right\}\leq 3\gamma_{\mathrm{alg},K}\max\left\{1,\gamma_{\mathrm{lin},K}\right\}\eta^{k,i}_{\mathrm{disc},K,\alpha}. \tag{80}$$

Employing now crucially assumption (77), it follows that

$$C_{\mathrm{st}}C_{\mathrm{cont,PF}}\sum_{\boldsymbol{a}\in\mathcal{V}_K}\left\|\mu_\alpha^{-\frac{1}{2}}\boldsymbol{\sigma}^{k,i}_{\alpha h,\mathrm{alg}}\right\|_{\omega_h^{\boldsymbol{a}}}\leq\frac{\eta^{k,i}_{\mathrm{disc},K,\alpha}}{2}. \tag{81}$$

Finally, we combine (81) with (79) to bound $\eta^{k,i}_{\mathrm{disc},K,\alpha}$ without the term containing $\boldsymbol{\sigma}^{k,i}_{\alpha h,\mathrm{alg}}$, and we conclude using (78).

If $p\geq 2$ the analogue of equation (78) reads

$$\eta^{k,i}_{\mathrm{disc},K,\alpha}+\eta^{k,i}_{\mathrm{alg},K,\alpha}\leq\frac{\delta_K}{2C_{\mathrm{st}}C_{\mathrm{cont,PF}}}\eta^{k,i}_{\mathrm{disc},K,\alpha}.$$

While, inequalities (79) and (81) remain the same, inequality (80) reads

$$\sum_{\boldsymbol{a}\in\mathcal{V}_K}\left\|\mu_\alpha^{-\frac{1}{2}}\boldsymbol{\sigma}^{k,i}_{\alpha h,\mathrm{alg}}\right\|_{\omega_h^{\boldsymbol{a}}}\leq 3\gamma_{\mathrm{alg},K}\eta^{k,i}_{\mathrm{disc},K,\alpha}. \tag{82}$$

The conclusion follows immediately. □

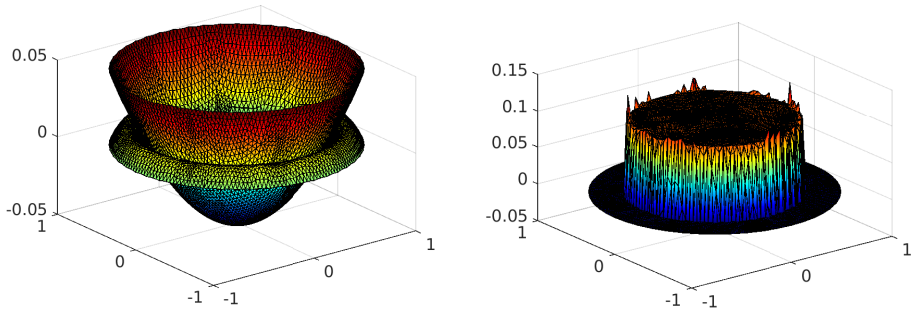

Figure 1: Solution at convergence for approximately 8000 mesh elements, ($p = 1$). Left: position of the membranes (u_{1h}, u_{2h}). Right: discrete action (λ_h).

8 Numerical experiments

This section illustrates numerically our theoretical developments in the case of continuous and piecewise affine and piecewise quadratic finite elements, $p = 1, 2$. We consider the unit disk $\Omega := \{(r, \theta) \in [0, 1] \times [0, 2\pi]\}$ using the polar coordinates, and an analytical solution given in [6] by, for all $(r, \theta) \in \Omega$,

$$u_1(r,\theta) := g(2r^2 - 1),$$
$$u_2(r,\theta) := \begin{cases} g(2r^2-1) & \text{if } r \le 1/\sqrt{2}, \\ g(1-r)(2r^2-1)\dfrac{\sqrt{2}}{\sqrt{2}-1} & \text{if } r \ge 1/\sqrt{2}, \end{cases} \quad \lambda(r,\theta) := \begin{cases} 2g & \text{if } r \le 1/\sqrt{2}, \\ 0 & \text{if } r \ge 1/\sqrt{2}. \end{cases}$$

This triple is the solution of the system (6) for the data f_1 and f_2 given by

$$f_1(r,\theta) := \begin{cases} -10g & \text{if } r \le 1/\sqrt{2}, \\ -8g & \text{if } r \ge 1/\sqrt{2}, \end{cases} \quad f_2(r,\theta) := \begin{cases} -6g & \text{if } r \le 1/\sqrt{2}, \\ -g\dfrac{1+8r-18r^2}{r}\dfrac{\sqrt{2}}{\sqrt{2}-1} & \text{if } r \ge 1/\sqrt{2}. \end{cases}$$

The parameters μ_1 and μ_2 are set to 1 and the boundary condition for the first membrane g is equal to 0.05.

We use the semismooth Newton linearization of Section 3.1 with the min function (27a) exemplified in Section 3.2, combined with the GMRES linear solver for the system (29) in the sense of Section 3.3. Figure 1 displays the behavior of the solution when the Newton-min and the GMRES solvers have converged. We observe a contact zone in the area $r \lesssim 1/\sqrt{2}$, where λ_h is positive. For the computation of $\boldsymbol{\sigma}^{k,i}_{\alpha h,\mathrm{alg}}$, $\alpha = 1, 2$, following Section 4.3, we consider a hierarchy with three uniformly refined meshes. To approximate the integrals containing $\lambda_h^{k,i,\mathrm{pos}}$ or $\lambda_h^{k,i,\mathrm{neg}}$, we use a Gauss quadrature formula with 7 ($p = 1$), respectively 16 ($p = 2$) points per element. Following (28) and (31), we define the linearization and algebraic residuals by

$$\boldsymbol{R}^{k,i}_{\mathrm{lin}} := \begin{pmatrix} \boldsymbol{F} - \mathbb{E}\boldsymbol{X}^{k,i}_h \\ -\mathbf{C}(\boldsymbol{X}^{k,i}_h) \end{pmatrix} \quad \text{and} \quad \boldsymbol{R}^{k,i}_{\mathrm{alg}} := \boldsymbol{R}^{k,i}_h. \tag{83}$$

Three different methods are tested:
1) The *exact* Newton-min method, where both the linear and nonlinear solvers are iterated to "almost" convergence: we set $\varepsilon_{\mathrm{alg}} := 2{\cdot}10^{-12}$ and $\varepsilon_{\mathrm{lin}} := 10^{-10}$, and use the criteria on the relative residuals

$$\text{(a)}\ \left\|\boldsymbol{R}^{k,i}_{\mathrm{alg}}\right\| / \left\|\boldsymbol{B}^{k-1}\right\| \le \varepsilon_{\mathrm{alg}}, \qquad \text{(b)}\ \left\|\boldsymbol{R}^{k,i}_{\mathrm{lin}}\right\| / \left\|\begin{pmatrix} \boldsymbol{F} \\ 0 \end{pmatrix}\right\| \le \varepsilon_{\mathrm{lin}}. \tag{84}$$

Thus $\boldsymbol{X}^{k,i}_h \approx \boldsymbol{X}_h$, where $\boldsymbol{X}_h$ is the solution of (28).
2) The *inexact* Newton-min method, where $\alpha_{\mathrm{alg}} := 1$, $\varepsilon_{\mathrm{lin}} := 10^{-10}$, and

$$\text{(a)}\ \left\|\boldsymbol{R}^{k,i}_{\mathrm{alg}}\right\| / \left\|\boldsymbol{B}^{k-1}\right\| \le \alpha_{\mathrm{alg}} \left\|\boldsymbol{R}^{k,i}_{\mathrm{lin}}\right\| / \left\|\begin{pmatrix} \boldsymbol{F} \\ 0 \end{pmatrix}\right\|, \quad \text{(b)}\ \left\|\boldsymbol{R}^{k,i}_{\mathrm{lin}}\right\| / \left\|\begin{pmatrix} \boldsymbol{F} \\ 0 \end{pmatrix}\right\| \le \varepsilon_{\mathrm{lin}}. \tag{85}$$

3) Our *adaptive inexact* Newton-min method of Algorithm 1, using the stopping criteria (63) with $\gamma_{\mathrm{alg}} := 0.3$ and $\gamma_{\mathrm{lin}} := 0.3$.

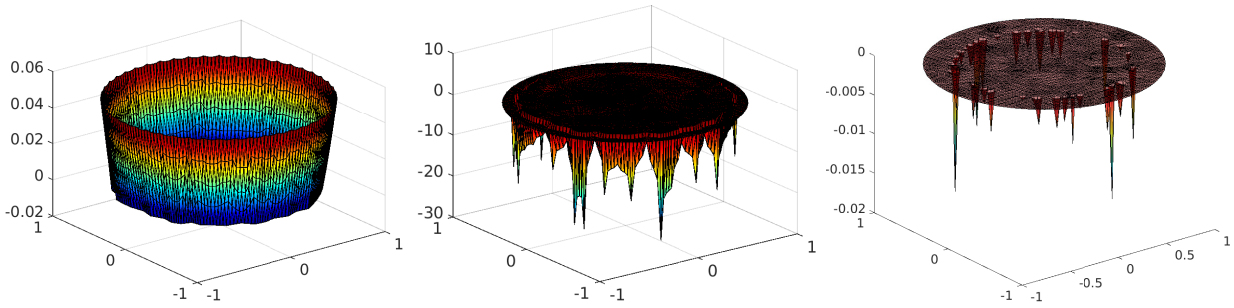

Figure 2: Left: $u_{1h}^{k,i} - u_{2h}^{k,i}$ at the second Newton-min step ($p = 1$, $k = 2$, $i = 20$). Center: $\lambda_h^{k,i}$ at the third Newton-min step ($p = 1$, $k = 3$, $i = 20$). Right: $\lambda_h^{\mathrm{neg}} := \min\{\lambda_h, 0\}$ at convergence ($p = 2$, $k = \overline{k}$, $i = \overline{i}$); in figures, $\mathbb{P}_2$ functions are plotted as $\mathbb{P}_1$ functions on a refined mesh (each triangle divided into 4 triangles).

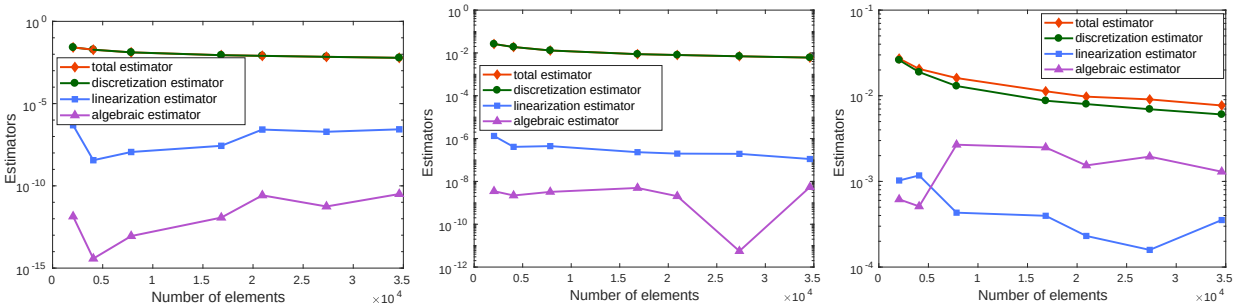

Figure 3: $p = 1$: a posteriori estimators $\eta^{\overline{k},\overline{i}}$, $\eta_{\mathrm{disc}}^{\overline{k},\overline{i}}$, $\eta_{\mathrm{lin}}^{\overline{k},\overline{i}}$, $\eta_{\mathrm{alg}}^{\overline{k},\overline{i}}$ at convergence as a function of the number of mesh elements. Exact (left), inexact (middle), and adaptive inexact (right) Newton-min methods with respectively the stopping criteria (84), (85), and (63). The log scales are different in each graph.

In the cases of inexact and adaptive inexact methods, the criteria are computed every $\nu := 10$ linear solver iterations. An ILU preconditioner is used to speed up the GMRES solver. The initial linearization guess is taken as $(\boldsymbol{X}_h^0)^T := [g\mathbf{1}, \mathbf{0}, \mathbf{0}]^T \in \mathbb{R}^{3\mathcal{N}^{p,\mathrm{i}}}$. In the sequel, when the stopping criterion of the nonlinear solver is satisfied, the index k will be denoted by $\overline{k}$, and similarly for the index i with $\overline{i}$. The results are presented for a mesh containing approximately 8000 triangles, except when looking at mesh dependency.

Figure 2 shows the possible violation of the physical constraints, see Remark 5.2. For piecewise affine finite elements, during the iterations before convergence, $u_{1h}^{k,i} < u_{2h}^{k,i}$ and $\lambda_h^{k,i} < 0$ can occur, see the left and center figures. Even at convergence, $u_{1h} < u_{2h}$ and $\lambda_h < 0$ can occur with piecewise quadratic elements, see the right figure for λ_h, where small undershoots take place.

8.1 Numerical results for piecewise affine elements ($p = 1$)

We first investigate the case $p = 1$. Figure 3 displays the curves of the different estimators as a function of the number of mesh elements when the nonlinear and algebraic stopping criteria (84), (85), or (63) are satisfied. In this example, the total estimators $\eta^{\overline{k},\overline{i}}$ (41) are almost identical for the three methods (exact, inexact, and adaptive inexact). Moreover, one observes that $\eta^{\overline{k},\overline{i}} \approx \eta_{\mathrm{disc}}^{\overline{k},\overline{i}}$, and the error components from Newton-min and GMRES are relatively small. Next, $\eta_{\mathrm{alg}}^{\overline{k},\overline{i}}$ takes values below 10^{-11} for the exact semismooth Newton and below 10^{-8} for the inexact semismooth Newton, whereas $\eta_{\mathrm{lin}}^{\overline{k},\overline{i}}$ takes similar values in both cases (below 10^{-6}). The adaptive inexact Newton method proposed here shows a different behavior: both $\eta_{\mathrm{alg}}^{\overline{k},\overline{i}}$ and $\eta_{\mathrm{lin}}^{\overline{k},\overline{i}}$ take larger values that are just sufficiently small not to influence the overall error estimator. It is also interesting to note the following: although the relative linearization residual $\|\boldsymbol{R}_{\mathrm{lin}}^{k,i}\|/\|\begin{pmatrix}\boldsymbol{F}\\ 0\end{pmatrix}\|$ is requested to lie below $\varepsilon_{\mathrm{lin}} = 10^{-10}$ in (84)(b) and (85)(b), our linearization estimator $\eta_{\mathrm{lin}}^{k,i}$ given by (62d) still remains quite large ($\approx 10^{-6}$, see Figure 3, left and middle). Clearly, the linearization residual and our linearization estimator expressing its lifting back to the physical space can have significantly different orders

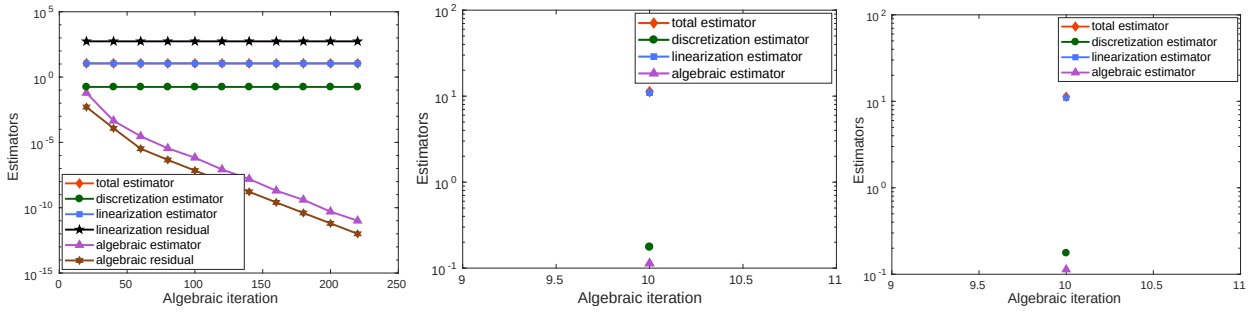

Figure 4: $p = 1$: estimators as a function of the algebraic iterations for $k = 1$. Exact (left), inexact (middle), and adaptive inexact (right) Newton-min methods.

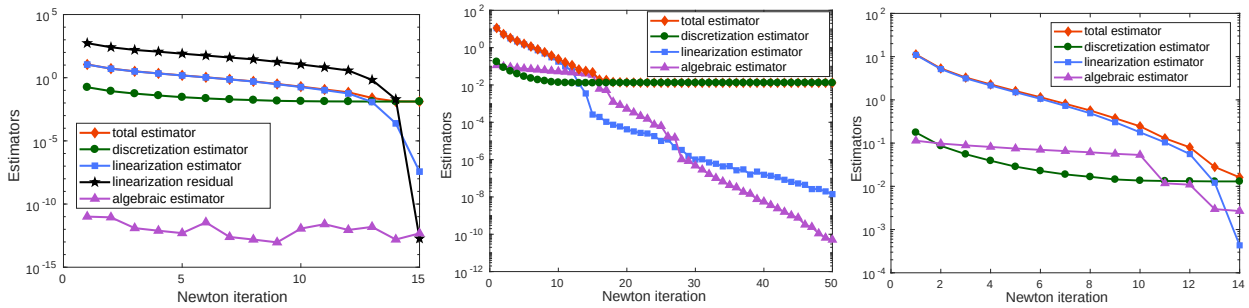

Figure 5: $p = 1$: estimators as a function of the Newton-min iterates k ($i = \bar{i}$). Exact (left), inexact (middle), and adaptive inexact (right) Newton-min methods.

of magnitude.

Figure 4 shows the evolution of the various estimators and of the (non-relative) residuals $\|\boldsymbol{R}_{\text{lin}}^{k,i}\|$ and $\|\boldsymbol{R}_{\text{alg}}^{k,i}\|$ during the algebraic iterations of the first Newton-min step ($k = 1$, i varies). In the exact case, we observe that 220 GMRES iterations are needed to achieve the criterion (84)(a). In both the inexact and adaptive inexact cases, only 10 GMRES iterations are required to satisfy respectively (85)(a) and (63)(a), the estimators are computed only once (recall $\nu = 10$), and the total and linearization estimators are approximately equal.

Figure 5 represents the evolution of the various estimators as a function of the semismooth Newton iterations when the algebraic solver stopping criteria have been satisfied (k varies, $i = \bar{i}$). For the three methods, the linearization estimator dominates and is close to the total estimator until approximately the 14^{th} iteration. Next, one can observe that during the Newton-min iterations, the linearization estimator steadily decreases, whereas the discretization one roughly stagnates. The linearization iterations are then stopped in the adaptive inexact Newton-min case when the discretization error becomes dominant, whereas the inexact Newton-min performs many unnecessary additional iterations. This can also be the case for the exact Newton-min algorithm in general, but here it converges very rapidly at the end. Criteria (84)(b) (exact), (85)(b) (inexact), and (63)(b) (adaptive inexact) are met respectively in 15, 46, and 14 iterations.

Figure 6 illustrates the overall performance of the three approaches. In the first graph, the behavior of the three methods is represented when the number of mesh elements is increased. The inexact Newton-min method requires many more semismooth iterations to converge in comparison with the other methods. The exact and the adaptive inexact methods lead to roughly the same number of nonlinear iterations. The second graph of Figure 6 presents the required number of algebraic steps to satisfy the linear stopping criterion for each method at each Newton-min step for a given mesh. Many algebraic iterations are necessary in the exact Newton-min case, while in the inexact and adaptive inexact cases, the algebraic solver is generally stopped in 10 iterations. The total number of algebraic iterations is displayed as a function of the number of elements in the right part of Figure 6. We observe that exact Newton-min is the most expensive method (3000 iterations for 35 000 elements), whereas inexact and adaptive inexact require respectively 1660 and 670 iterations. Thus, globally our approach yields an economy by a factor of roughly 2 with respect to

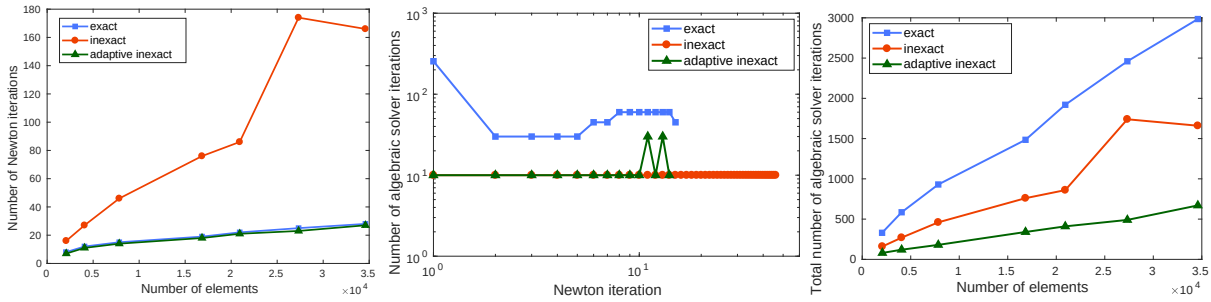

Figure 6: $p = 1$: number of Newton-min iterations per number of elements (left), number of algebraic solver iterations per Newton-min step for 8000 elements (middle), and total number of linear solver iterations per number of elements (right).

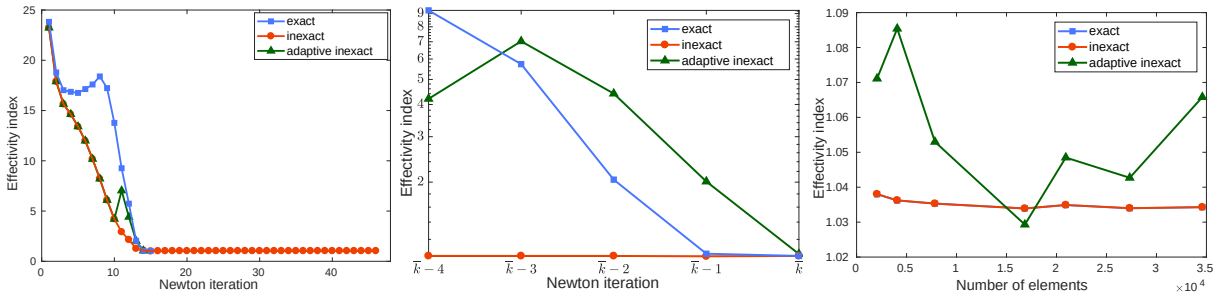

Figure 7: $p = 1$: effectivity index as a function of the Newton-min steps for three methods (left), zoom for the last five iterates (middle), and effectivity indices as a function of the number of mesh elements (right); $\overline{k}$ stands for the last Newton-min step for each method ($\overline{k} = 15$, 46, and 14 respectively for exact, inexact and adaptive inexact methods).

inexact Newton-min and roughly 5 with respect to exact Newton-min in terms of total algebraic solver iterations.

The effectivity indices, defined as the ratio of the total estimator $\eta^{k,\overline{i}}$ over the energy norm $|||\boldsymbol{u} - \boldsymbol{u}_h^{k,\overline{i}}|||$, are displayed in Figure 7 as a function of the Newton-min iterations for the three methods (k varies, $i = \overline{i}$.) We observe that they always decrease to the optimal value 1 when the computational effort grows. In the middle part of Figure 7, we zoom on the last five semismooth Newton iterations for all the methods. In the right part of Figure 7, we displayed the value of the effectivity indices for each method for several number of mesh elements when the Newton-min solver and the GMRES solver have converged ($k = \overline{k}$, $i = \overline{i}$). Note that the curves of inexact and adaptive inexact Newton-min are superimposed. We observe that increasing the mesh size will not influence the behavior of the effectivity indices. It is indeed still close to the optimal value of 1.

Figure 8 shows the local distribution of the total error estimator $\eta^{k,\overline{i}}$ and of the error in the energy norm $|||\boldsymbol{u} - \boldsymbol{u}_h^{k,\overline{i}}|||$ for the adaptive inexact Newton-min method ($k = 3$, $i = \overline{i}$). We observe a very close agreement, even in the presence of algebraic and linearization errors.

Finally, Table 1 shows the dependency of our adaptive inexact method on the coefficients γ_{lin} and γ_{alg} in

Table 1: Number of iterations for the adaptive inexact Newton-min method for several parameters γ_{alg} and γ_{lin}.

$(\gamma_{\text{alg}}, \gamma_{\text{lin}})$	$(0.3, 0.3)$	$(0.03, 0.3)$	$(0.3, 0.03)$	$(0.03, 0.03)$
Newton-min iterations	26	26	27	27
Average algebraic iterations	26	43	25	42
Total iterations	670	1130	680	1140

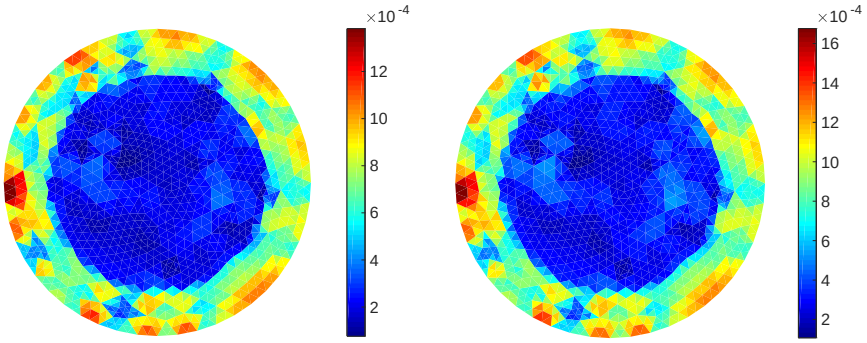

Figure 8: Error in energy norm (left) and total estimator (right), adaptive inexact Newton-min method, $p = 1$, 8000 elements.

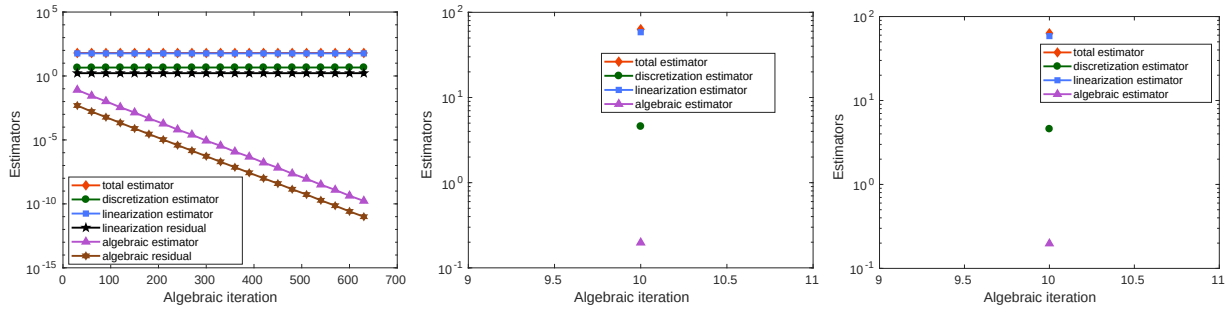

Figure 9: $p = 2$: estimators as a function of the algebraic iterations for $k = 1$. Exact (left), inexact (middle), and adaptive inexact (right) Newton-min methods.

the algebraic and linearization stopping criteria (63)(a) and (63)(b) (on the finest mesh with 35000 elements). The first line gives the number of Newton-min iterations required to satisfy (63)(b), and the second one the number of algebraic iterations required to meet (63)(a), averaged over all Newton-min iterations. As the linearization convergence is fast, the choice of γ_{lin} has a very small impact, but choosing γ_{alg} small adds many additional iterations. In any case, however, the overall number of algebraic iterations remains (much) smaller than for the exact and inexact semismooth Newton methods.

8.2 Numerical results for piecewise quadratic elements ($p = 2$)

Figures 9, 10, and 11 are respectively the counterparts of Figures 4, 5, and 6 for piecewise quadratic elements ($p = 2$). In this context, there are 4 times more degrees of freedom than in the case $p = 1$, and the discretization and linearization estimators are more intricate, see (62e)–(62f). The comments made in Section 8.1 remain globally valid. This gives us a good confidence in the identification of the various components of the error also for $p = 2$, see Corollary 5.6.

Here, the costs of the exact and inexact Newton-min methods have importantly increased: in both cases, when $p = 2$, the number of Newton-min iterations has more than doubled. The total number of algebraic iterates has been multiplied by a factor of roughly 9 (21 000 iterates instead of 2400 for $p = 1$ in the exact case for a mesh with 27 000 elements), or roughly 4 (7300 iterates instead of 1700 in the inexact case). In contrast, the adaptive inexact method remains cheap in terms of Newton-min and algebraic iterations: only 540 total algebraic iterates are required for the 27 000 element mesh, instead of 490. Figure 12 shows the effectivity indices for the three methods. They tend to values close to the optimal value of 1. The final "bumps" could not be explained, but they remain small.

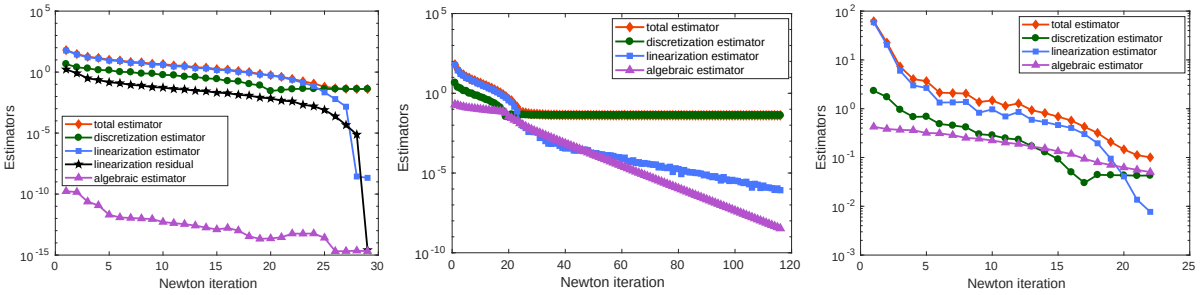

Figure 10: $p = 2$: estimators as a function of the Newton-min iterates k ($i = \overline{i}$). Exact (left), inexact (middle), and adaptive inexact (right) Newton-min methods.

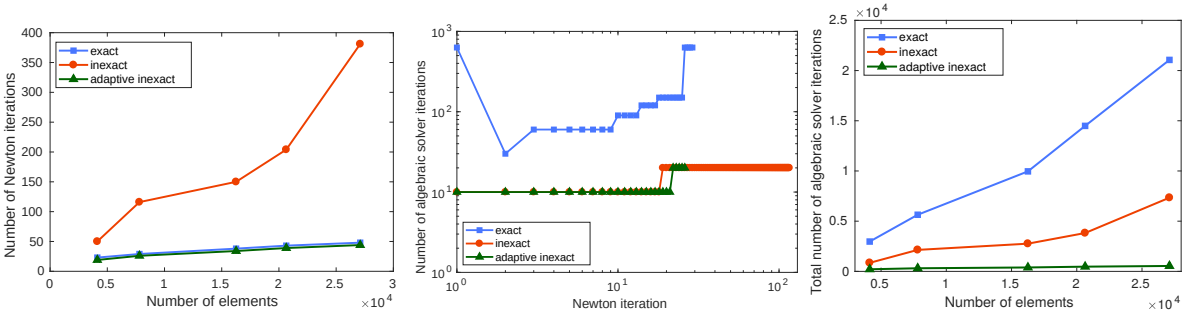

Figure 11: $p = 2$: number of Newton-min iterations per number of elements (left), number of algebraic solver iterations per Newton-min step for 8000 elements (middle), and total number of linear solver iterations per number of elements (right).

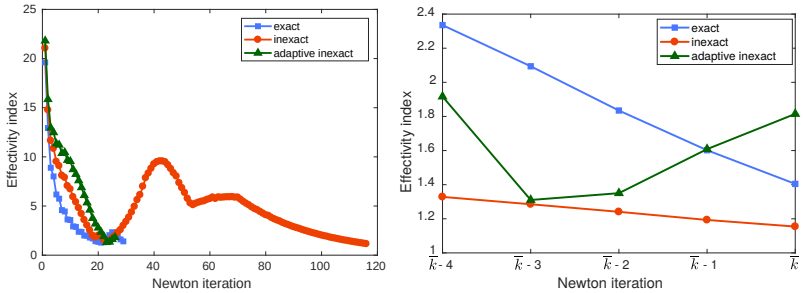

Figure 12: $p = 2$: effectivity index as a function of the Newton-min steps for three methods (left), zoom for the last five iterates (right); $\overline{k}$ stands for the last Newton-min step for each method ($\overline{k} = 29$, 116, and 22 respectively for exact, inexact and adaptive inexact methods).

9 Conclusions

In this work, we have designed an adaptive inexact semismooth Newton method with adaptive stopping criteria for the problem of contact between two membranes. We proved an optimal a posteriori error estimate between the exact and approximate solution on each semismooth Newton step $k \geq 1$ and on each algebraic solver step $i \geq 1$, for any polynomial degree p. This estimate enables to distinguish the different error components. Our numerical experiments for $p = 1, 2$ confirm that the adaptive inexact Newton-min method is much faster in comparison with the exact and inexact Newton-min ones. Moreover, in contrast to these standard methods, the adaptive inexact method presented here provides an accurate estimation of the error between the exact solution and its approximation. The extension of our developments to parabolic variational inequalities is addressed in [19].

References

[1] M. Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, Pure and Applied Mathematics (New York), Wiley-Interscience [John Wiley & Sons], New York, 2000, https://doi.org/10.1002/9781118032824.

[2] M. Ainsworth, J. T. Oden, and C.-Y. Lee, *Local a posteriori error estimators for variational inequalities*, Numer. Methods Partial Differential Equations, 9 (1993), pp. 23–33, https://doi.org/10.1002/num.1690090104.

[3] R. Becker, C. Johnson, and R. Rannacher, *Adaptive error control for multigrid finite element methods*, Computing, 55 (1995), pp. 271–288, https://doi.org/10.1007/BF02238483.

[4] F. Ben Belgacem, C. Bernardi, A. Blouza, and M. Vohralík, *A finite element discretization of the contact between two membranes*, M2AN Math. Model. Numer. Anal., 43 (2008), pp. 33–52, https://doi.org/10.1051/m2an/2008041.

[5] F. Ben Belgacem, C. Bernardi, A. Blouza, and M. Vohralík, *On the unilateral contact between membranes. Part 1: Finite element discretization and mixed reformulation*, Math. Model. Nat. Phenom., 4 (2009), pp. 21–43, https://doi.org/10.1051/mmnp/20094102.

[6] F. Ben Belgacem, C. Bernardi, A. Blouza, and M. Vohralík, *On the unilateral contact between membranes. Part 2: a posteriori analysis and numerical experiments*, IMA J. Numer. Anal., 32 (2012), pp. 1147–1172, https://doi.org/10.1093/imanum/drr003.

[7] I. Ben Gharbia and J. C. Gilbert, *An algorithmic characterization of* $\mathbf{P}$*-matricity*, SIAM J. Matrix Anal. Appl., 34 (2013), pp. 904–916.

[8] J. F. Bonnans, J. C. Gilbert, C. Lemaréchal, and C. A. Sagastizábal, *Numerical optimization*, Universitext, Springer-Verlag, Berlin, second ed., 2006. Theoretical and practical aspects.

[9] D. Braess, *A posteriori error estimators for obstacle problems—another look*, Numer. Math., 101 (2005), pp. 415–421, https://doi.org/10.1007/s00211-005-0634-1.

[10] D. Braess, V. Pillwein, and J. Schöberl, *Equilibrated residual error estimates are p-robust*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 1189–1197, https://doi.org/10.1016/j.cma.2008.12.010.

[11] D. Braess and J. Schöberl, *Equilibrated residual error estimator for edge elements*, Math. Comp., 77 (2008), pp. 651–672, https://doi.org/10.1090/S0025-5718-07-02080-7.

[12] F. Brezzi and M. Fortin, *Mixed and hybrid finite element methods*, vol. 15 of Springer Series in Computational Mathematics, Springer-Verlag, New York, 1991, https://doi.org/10.1007/978-1-4612-3172-1.

[13] F. Brezzi, W. W. Hager, and P.-A. Raviart, *Error estimates for the finite element solution of variational inequalities*, Numer. Math., 28 (1977), pp. 431–443, https://doi.org/10.1007/BF01404345.

[14] P. N. Brown and Y. Saad, *Hybrid Krylov methods for nonlinear systems of equations*, SIAM J. Sci. Statist. Comput., 11 (1990), pp. 450–481, https://doi.org/10.1137/0911026.

[15] M. Bürg and A. Schröder, *A posteriori error control of hp-finite elements for variational inequalities of the first and second kind*, Comput. Math. Appl., 70 (2015), pp. 2783–2802, https://doi.org/10.1016/j.camwa.2015.08.031.

[16] Z. Chen and R. H. Nochetto, *Residual type a posteriori error estimates for elliptic obstacle problems*, Numer. Math., 84 (2000), pp. 527–548, https://doi.org/10.1007/s002110050009.

[17] F. H. Clarke, *Optimization and nonsmooth analysis*, vol. 5 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second ed., 1990, https://doi.org/10.1137/1.9781611971309.

[18] J. Dabaghi, *A posteriori error estimates for variational inequalities: application to a two-phase flow in porous media*, Ph.D. thesis, Sorbonne Université, June 2019, https://hal.archives-ouvertes.fr/tel-02151951.

[19] J. Dabaghi, V. Martin, and M. Vohralík, *A posteriori estimates distinguishing the error components and adaptive stopping criteria for numerical approximations of parabolic variational inequalities*. HAL Preprint 02274493, submitted, 2019, https://hal.archives-ouvertes.fr/hal-02274493.

[20] T. De Luca, F. Facchinei, and C. Kanzow, *A semismooth equation approach to the solution of nonlinear complementarity problems*, Math. Programming, 75 (1996), pp. 407–439.

[21] P. Destuynder and B. Métivet, *Explicit error bounds in a conforming finite element method*, Math. Comp., 68 (1999), pp. 1379–1396, https://doi.org/10.1090/S0025-5718-99-01093-5.

[22] S. C. Eisenstat and H. F. Walker, *Globally convergent inexact Newton methods*, SIAM J. Optim., 4 (1994), pp. 393–422, https://doi.org/10.1137/0804022.

[23] A. Ern and M. Vohralík, *Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs*, SIAM J. Sci. Comput., 35 (2013), pp. A1761–A1791, https://doi.org/10.1137/120896918.

[24] A. Ern and M. Vohralík, *Polynomial-degree-robust a posteriori estimates in a unified setting for conforming, nonconforming, discontinuous Galerkin, and mixed discretizations*, SIAM J. Numer. Anal., 53 (2015), pp. 1058–1081, https://doi.org/10.1137/130950100.

[25] F. Facchinei and C. Kanzow, *A nonsmooth inexact Newton method for the solution of large-scale nonlinear complementarity problems*, Math. Programming, 76 (1997), pp. 493–512.

[26] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems. Vol. I*, Springer Series in Operations Research, Springer-Verlag, New York, 2003.

[27] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems. Vol. II*, Springer Series in Operations Research, Springer-Verlag, New York, 2003.

[28] Z. Ge, Q. Ni, and X. Zhang, *A smoothing inexact Newton method for variational inequalities with nonlinear constraints*, J. Inequal. Appl., (2017), pp. Paper No. 160, 12.

[29] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2002), pp. 865–888 (2003), https://doi.org/10.1137/S1052623401383558.

[30] I. Hlaváček, J. Haslinger, J. Nečas, and J. Lovíšek, *Solution of variational inequalities in mechanics*, vol. 66 of Applied Mathematical Sciences, Springer-Verlag, New York, 1988, https://doi.org/10.1007/978-1-4612-1048-1. Translated from the Slovak by J. Jarník.

[31] S. Hüeber, G. Stadler, and B. I. Wohlmuth, *A primal-dual active set algorithm for three-dimensional contact problems with Coulomb friction*, SIAM J. Sci. Comput., 30 (2008), pp. 572–596, https://doi.org/10.1137/060671061.

[32] K. Ito and K. Kunisch, *Semi-smooth Newton methods for variational inequalities of the first kind*, M2AN Math. Model. Numer. Anal., 37 (2003), pp. 41–62, https://doi.org/10.1051/m2an:2003021.

[33] K. Ito and K. Kunisch, *Lagrange multiplier approach to variational problems and applications*, vol. 15 of Advances in Design and Control, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008, https://doi.org/10.1137/1.9780898718614.

[34] K. Ito and K. Kunisch, *Semi-smooth Newton methods for the Signorini problem*, Appl. Math., 53 (2008), pp. 455–468, https://doi.org/10.1007/s10492-008-0036-7.

[35] P. Jiránek, Z. Strakoš, and M. Vohralík, *A posteriori error estimates including algebraic error and stopping criteria for iterative solvers*, SIAM J. Sci. Comput., 32 (2010), pp. 1567–1590, https://doi.org/10.1137/08073706X.

[36] C. Kanzow, *An active set-type Newton method for constrained nonlinear systems*, in Complementarity: applications, algorithms and extensions (Madison, WI, 1999), vol. 50 of Appl. Optim., Kluwer Acad. Publ., Dordrecht, 2001, pp. 179–200, https://doi.org/10.1007/978-1-4757-3279-5_9.

[37] C. Kanzow, *Inexact semismooth Newton methods for large-scale complementarity problems*, Optim. Methods Softw., 19 (2004), pp. 309–325. The First International Conference on Optimization Methods and Software. Part II.

[38] C. T. Kelley, *Iterative methods for linear and nonlinear equations*, vol. 16 of Frontiers in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1995. With separately available software.

[39] R. Kornhuber, *A posteriori error estimates for elliptic variational inequalities*, Comput. Math. Appl., 31 (1996), pp. 49–60.

[40] K. Kunisch and G. Stadler, *Generalized Newton methods for the 2D-Signorini contact problem with friction in function space*, M2AN Math. Model. Numer. Anal., 39 (2005), pp. 827–854, https://doi.org/10.1051/m2an:2005036.

[41] J.-L. Lions and G. Stampacchia, *Variational inequalities*, Comm. Pure Appl. Math., 20 (1967), pp. 493–519.

[42] F. Louf, J.-P. Combe, and J.-P. Pelle, *Constitutive error estimator for the control of contact problems involving friction*, Comput. & Structures, 81 (2003), pp. 1759–1772, https://doi.org/10.1016/S0045-7949(03)00200-1.

[43] J. M. Martínez and L. Q. Qi, *Inexact Newton methods for solving nonsmooth equations*, J. Comput. Appl. Math., 60 (1995), pp. 127–145. Linear/nonlinear iterative methods and verification of solution (Matsuyama, 1993).

[44] J. Papež, U. Rüde, M. Vohralík, and B. Wohlmuth, *Sharp algebraic and total a posteriori error bounds for h and p finite elements via a multilevel approach. Recovering mass balance in any situation.* HAL Preprint 01662944, submitted for publication, 2019, https://hal.inria.fr/hal-01662944/.

[45] J. Papež, Z. Strakoš, and M. Vohralík, *Estimating and localizing the algebraic and total numerical errors using flux reconstructions*, Numer. Math., 138 (2018), pp. 681–721, https://doi.org/10.1007/s00211-017-0915-5.

[46] S. Repin, *A posteriori estimates for partial differential equations*, vol. 4 of Radon Series on Computational and Applied Mathematics, Walter de Gruyter GmbH & Co. KG, Berlin, 2008, https://doi.org/10.1515/9783110203042.

[47] S. I. Repin, *Functional a posteriori estimates for elliptic variational inequalities*, Zap. Nauchn. Sem. S.-Peterburg. Otdel. Mat. Inst. Steklov. (POMI), 348 (2007), pp. 147–164, 305, https://doi.org/10.1007/s10958-008-9093-4.

[48] G. Stadler, *Semismooth Newton and augmented Lagrangian methods for a simplified friction problem*, SIAM J. Optim., 15 (2004), pp. 39–62, https://doi.org/10.1137/S1052623403420833.

[49] G. Stadler, *Path-following and augmented Lagrangian methods for contact problems in linear elasticity*, J. Comput. Appl. Math., 203 (2007), pp. 533–547, https://doi.org/10.1016/j.cam.2006.04.017.

[50] M. Ulbrich, *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces*, vol. 11 of MOS-SIAM Series on Optimization, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2011, https://doi.org/10.1137/1.9781611970692.

[51] M. Ulbrich, S. Ulbrich, and D. Bratzke, *A multigrid semismooth Newton method for semilinear contact problems*, J. Comput. Math., 35 (2017), pp. 486–528, https://doi.org/10.4208/jcm.1702-m2016-0679.

[52] A. Veeser, *Efficient and reliable a posteriori error estimators for elliptic obstacle problems*, SIAM J. Numer. Anal., 39 (2001), pp. 146–167, https://doi.org/10.1137/S0036142900370812.

[53] R. Verfürth, *A posteriori error estimation techniques for finite element methods*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2013, https://doi.org/10.1093/acprof:oso/9780199679423.001.0001.

[54] S. J. Wright, *Primal-dual interior-point methods*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997, https://doi.org/10.1137/1.9781611971453.

[55] S. Zhang, Y. Yan, and R. Ran, *Path-following and semismooth Newton methods for the variational inequality arising from two membranes problem*, J. Inequal. Appl., (2019), pp. Paper No. 1, 13, https://doi.org/10.1186/s13660-019-1955-4.